ESD-TR-65-79 TM-03728

# A COMPUTERIZED TECHNIQUE TO EXPRESS UNCERTAINTY IN ADVANCED SYSTEM COST ESTIMATES

TECHNICAL REPORT NO. ESD-TR-65-79

NOVEMBER 1965

S. Sobel

Prepared for

COMPTROLLER (PROGRAMS DIVISION)
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 850

Prepared by

THE MITRE CORPORATION
Bedford, Massachusetts

Contract AF 19(628)-2390

ESD-TR-65-79 TM-03728

# A COMPUTERIZED TECHNIQUE TO EXPRESS UNCERTAINTY IN ADVANCED SYSTEM COST ESTIMATES

TECHNICAL REPORT NO. ESD-TR-65-79

NOVEMBER 1965

S. Sobel

Prepared for

COMPTROLLER (PROGRAMS DIVISION)
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 850

Prepared by

THE MITRE CORPORATION
Bedford, Massachusetts

Contract AF 19(628)-2390

## ABSTRACT

The technique described presents a method to express uncertainty quantitatively in advanced system cost estimates. In particular, the technique suggests the employment of subjective probability distributions, which describe the uncertainty in each system element, to determine an approximate distribution for total system cost. A 7090 program has been written to perform the computational operations.

## REVIEW AND APPROVAL

This technical report has been reviewed and is approved.

Francis J. Hoermann

FRANCIS J. HOERMANN
Colonel, USAF
Comptroller

## TABLE OF CONTENTS



### APPENDICES

TABLE OF CONTENTS (cont'd.)



LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (cont'd.)

## GLOSSARY*

**Stochastic:** An adjective referring to an operation which describes uncertainty statistically.

**Random Variable:** A quantity whose value is uncertain. To fully describe a random variable, the probability of every possible value of the random variable must be given.

**Upper (Lower) Tail:** The value of a random variable such that the probability of this value or any higher (lower) value equals a stated amount.

**Mode:** The most probable value of a random variable.

**Dependency:** If $x$ and $y$ are dependent random variables, knowledge of the value of either variable will change the analyst's feeling about the distribution of the other. Mathematically, if $x$ and $y$ are dependent, then $p(x_k/y_j)$ can not equal $p(x_k)$ for all values of $k$ and $j$.

**Linearly Scaled Beta Distribution:**** A continuous finite unimodal distribution that may be either skew or symmetrical and has four degrees of freedom. The rectangular distribution is a special case of the scaled Beta, and the normal and gamma distributions are limiting forms.

---

*Definitions correspond to word use in this report.

**This distribution is equivalent to Pearson Type I.

Mathematical Expression:

$$P_{BETA}(x; \alpha, \beta, a, b) = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}\left(\frac{x-a}{b}\right)^{\alpha}\left(1-\frac{x-a}{b}\right)^{\beta},$$

where

$a$, $b$, $\alpha$, and $\beta$ are real numbers;

$b$, $\alpha$, and $\beta$ are non-negative;

$\Gamma(u)$ is the gamma function; and

$a \leq x \leq a + b$.

$k^{th}$ Origin Moment: The average (or expected value) of a random variable raised to the $k^{th}$ power. The first origin moment is the mean, $\mu$.

Mathematical Expression:

$$a^{(k)} = \underbrace{\sum_i x_i^k p_i}_{\text{discrete case}} \quad \text{or} \quad \underbrace{\int_{-\infty}^{+\infty} x^k p(x)\, dx}_{\text{continuous case}} .$$

$k^{th}$ Central Moment: The average (or expected value) of a random variable diminished by its mean raised to the $k^{th}$ power. The first central moment is zero and the second is the variance (the standard deviation squared).

Mathematical Expression:

$$\mu^{(k)} = \underbrace{\sum_i (x_i - \mu)^k p_i}_{\text{discrete case}} \quad \text{or} \quad \underbrace{\int_{-\infty}^{+\infty} (x-\mu)^k p(x)\, dx}_{\text{continuous case}} .$$

Additive
Moments: Moments (of random variables) which have the following property:* the $k^{th}$ moment for the sum of independent random variables equals the sum of the $k^{th}$ moments of each of the added variables.

Mathematical Expression for the First Four Additive Moments:

$$A_1 = \alpha^{(1)}, \; A_2 = \mu^{(2)}, \; A_3 = \mu^{(3)}, \; A_4 = \mu^{(4)} - 3\left[\mu^{(2)}\right]^2 \quad .$$

Multiplicative
Moments: Moments (of random variables) which have the following property:* the $k^{th}$ moment for the product of independent random variables equals the product of $k^{th}$ moments of each of the multiplied variables.

Mathematical Expression for Multiplicative Moments:

$$M_k = \alpha^{(k)} \quad .$$

System Cost
Input Elements: Each variable the analyst defines in the process of estimating the system total cost. Examples are quantities of personnel, equipment packages, and computer programs, price levels, and planning factors. As a convenience the abbreviated form, cost element, is used in the paper.

---

*See Appendix III for proof.

**Cost Structure:** The manner in which input cost elements are combined to determine total system cost.

## AUTHOR'S PREFACE

Although there is wide agreement concerning the logico-mathematical rules to which probabilities (whatever they may be) must adhere, there is substantial disagreement among applied mathematicians, decision theorists, and statisticians as to the essential meaning of probability. Currently, this controversy seems to be centered around two distinct points of view. One is the "objectivist" or "relative frequency" point of view, which defines probability as the long-run relative frequency limit of the ratio of the observed number of favorable events to the total number of observed instances associated with the outcomes of a random physical process. Probability, according to this point of view, is a phenomenological concept--a statistic estimated from repeated observations of some directly observable phenomenon. On the other hand, the "subjectivist" or "personalistic" point of view defines probability as a numerical coefficient purporting to measure a particular human being's subjective belief about the outcomes of some physical phenomenon. Probability, from the subjectivist standpoint, is not a phenomenological concept at all. It is not a characteristic of the physical phenomenon to which it purports to refer, but, rather, a characteristic of human beings--a component part of a particular individual's attitude toward a physical phenomenon. As such, subjective probability is a fictional concept, much like the concept of "force" in physics, which can only be inferred from observations of displacement and motion on the part of physical bodies, and like the concept of "intelligence" in psychology, which can only be inferred from observed verbal and/or choice behavior on the part of human beings.

Viewing probability from the objectivist point of view has proved highly successful in many historical applications, including the management of gambling houses, industrial quality control, and numerous scientific endeavors, particularly in the field of genetics. However, the success of these applications rests upon the existence of a stable, physical process of which repeated observations can be made. Even more, the whole notion of objective probability requires the existence of such processes, since the objectivist definition does not apply to a nonrepetitive phenomenon. Strictly interpreted, therefore, the objectivist point of view can be of no assistance in explaining or predicting the outcome(s) of nonrepetitive phenomena.

However, many real-world decision makers face the necessity of taking definite actions in the face of substantial uncertainty regarding the

outcomes of nonrepetitive phenomena. Since the objectivist concept of probability cannot be of service in making such decisions, the subjectivist view has been developed to fill the void. The conceptual foundations of subjective or personalistic probability have been worked out by Leonard J. Savage. [1]* Two pioneers in applying this concept to many real-world decisions problems have been Robert Schlaiffer [2] and Howard Raiffa. [3] Specific implementation of this concept in oil well drilling decisions is reported by C. Jackson Grayson. [4]

The position taken in this report is that the process by which military system costs are generated does not usually constitute a stable, repetitive phenomenon to which objectivist probability is an applicable concept. Therefore, probability is considered from the subjectivist point of view, and the author presents a technique by which the necessary implications of a decision-maker's or analyst's subjective judgments concerning costs may be derived and displayed.

J. R. Miller, III

*Numbers in brackets refer to references at end of report.

## SECTION I

## INTRODUCTION

### PURPOSE

The analyst faced with the task of estimating the cost of advanced military systems is continually beset by the problem of coping with a host of attendant uncertainties. Single-value estimates, although duly qualified, belie the full range of possible cost consequences that may ensue once the decision to acquire a given system has been made, and in a great many instances such estimates gravely misstate the ultimate system cost. Therefore, early in the system life cycle, defense planners have been promulgating their need for improved information about a proposed system's possible economic implications, despite manifest deficiencies in any early system description. The present paper suggests the kinds of information about the uncertainty which might be shown to defense planners and offers a technique to generate such information.

### SOURCES OF UNCERTAINTY

Uncertainties in advanced systems cost estimates arise from many sources. Examination of such sources has revealed that they may be divided into two basic categories. The first category, examples of which are shown in Table I, consists of those types of uncertainty which are amenable to planning decisions, that is, those types over which decision makers exercise control. The second category, examples of which are shown in Table II, is the class of uncertainties which is beyond direct control or explicit planning during the conceptual phase of a system.

## EXPRESSING COST UNCERTAINTY

Because of fundamental differences between the two categories of uncertainty, different techniques are needed in dealing with them. Category I uncertainties might best be represented by exhaustively identifying all significant competing design alternatives and by providing the decision maker with the economic implications of each planning choice. Nevertheless, because of the large lead time between the generation of advanced system cost estimates and the date by which the system achieves operational status, a large number of residual uncertainties, i.e., Category II uncertainties, will, in general, characterize each of the identified system alternatives. It would thus be desirable to convey to the decision maker a meaningful numerical range describing the magnitude of the effect on the total system cost of the simultaneous presence of all the residual uncertainties in the system elements. In addition, numerical ranges describing lower levels of cost aggregations than the total system are valuable to expose major sources of uncertainty. Finally, to support statistical analysis, it would be useful to show the relative likelihood that different cost magnitudes will be the actual ultimate total system cost. This report is addressed to generating this information.

Table I

Examples of Sources of Category I Uncertainty

| |
|---|
| System Performance Requirements* |
| Examples: coverage, numbers of users served, survivability, reliability, flexibility, growth potential, reaction time, range, warning time. |

*Although these requirements depend on factors (such as the enemy threat and existing technology) which a decision maker does not control, he does generate an explicit and conscious statement of the performance requirements for the system on the basis of a given set of constraints and goals; thus system performance requirements are included in Category I uncertainties.

Table I (cont'd.)

| |
|---|
| <u>Implementation Plan</u><br>Examples: schedules, amount of in-house vs. external management, number and extent of competing contractor studies during the program definition phase. |
| <u>Configuration (Resource Inventory)</u><br>Examples: number and kinds of hardware packages, the nature and extent of software, and the quantity and quality of personnel. ** |
| <u>Operational Concept</u><br>Examples: modes of operation during varying periods (e. g., tranquil, crisis, trans and post attack), amount of peace time exercises. |
| <u>Maintenance Concept</u><br>Examples: contractor vs. military maintenance, number of shifts. |
| <u>Logistics Concept</u><br>Examples: number, size, and locations of warehouses to maintain supply line, quantities and types of stocks and spares to be stored at each. |
| <u>Training Concept</u><br>Examples: in-house vs. contractor training, formal vs. on-the-job training. |
| <u>Funding Schedule</u><br>Examples: scheduling of obligational authority. |

**Ordinarily in the conceptual phase, these can be only grossly defined.

Table II

Examples of Sources of Category II Uncertainty

| |
|---|
| Price levels for equipment and contractor services |
| Extent of R&D effort needed for desired equipment development and resulting production cost of developed items |
| Exact quantities and types of personnel required to effectuate a given operations and maintenance concept |
| Exact quantities and types of support equipment (AGE), initial and follow-on spares, documentation, etc. |
| Manpower and computer time requirements for system design and management |
| Exact hardware and software design specifications for operational system |

## A FEW EARLY APPROACHES TO DESCRIBE CATEGORY II UNCERTAINTIES

A common practice in cost estimating is to study the nature and magnitude of the uncertainty in each of the system cost input elements* (see Glossary) and footnote the estimate with statements about possible variations in their values. This procedure serves to put the user of a cost estimate on notice that the single value shown for total system cost is subject to error, and helps to focus attention on the major sources of uncertainty. Nevertheless, such information does not reveal the extent to which the estimated total system cost is likely to deviate from the actual cost.

In order to provide some estimate of uncertainty in total system cost, analysts have attempted to make an astute guess about the variability of such

*For simplicity, the abbreviated form, cost element, will be used subsequently throughout this report.

cost and record the result for the decision maker's reference. The analyst performs this operation by mentally assessing the combined impact of all sources of uncertainty in all of the cost elements. For a typical size system, such practice strains the faculties of even the most gifted analyst, and hence the results are of questionable validity.

Another expedient used in quest of more reliable information about the variability in total costs requires the analyst to specify a likely range for each of the cost elements; then two additional total system cost estimates can be computed, one being the sum of the lowest values for all the cost elements, and the other being the sum of the highest values. In this manner a range is established for the total system cost. However, such a range is subject to strong criticism. It is self-evident that the possibility that <u>all</u> elements will actually attain their lowest values (or their highest) is very remote. Hence, such a range is a serious overstatement of the magnitude of likely variability in total system costs.

A more precise technique is therefore needed to translate the analyst's feelings about the uncertainty in all the cost elements into a statement about the uncertainty in the total system cost. This paper has been written to suggest one such technique.

## SECTION II

## DESCRIPTION OF THE PROPOSED TECHNIQUE

### GENERAL

Uncertainty in each of the cost elements is described by treating the elements as random variables. A subjective distribution[2] for each element is provided by the analyst to represent his feelings about the relative likelihood of all feasible values. The burden placed on the analyst, however, is minimized by requesting that he provide only four numbers which are estimates of four parameters characterizing the shape of the distribution curve.

From the four parameters the entire distribution is inferred. This is done by assuming a linearly scaled beta function (see Glossary) which satisfies the analyst's inputs. The element distributions are then combined in accordance with a cost structure[5] to produce the probability distribution for the total system cost and equivalent distributions for intermediate levels of cost aggregation. Cumulative probability intervals and relevant summary statistics (e.g., mode, mean, or median) can be generated from such distributions.

### LIMITATIONS

The procedure described is suited for direct application in the following situation: (1) the probability distributions specified in the input data are unimodal; (2) only additive and multiplicative operations are involved in relating element costs to total costs; (3) elements which are multiplied together are independent; and (4) elements which are added together are either independent or linearly dependent (see Appendix I).

Procedures are presented for permitting application of the technique in some situations which do not satisfy the above requirements.

## INPUTS PROVIDED BY THE COST ANALYST

To cope with uncertainty in schedule estimates, PERT[6] has adopted a procedure which utilizes a scaled beta function and requires that the analyst furnish three numbers to specify its form. These three numbers are estimates of the lowest, $X_L$ , and highest, $X_H$ , values (lower and upper 1 percent tails, respectively) and the mode, $X_P$ , (most probable value).

The PERT procedure assumes that the distribution's standard deviation can always be satisfactorily approximated as one-sixth of the total range, i.e., $1/6 \; X_H - X_L$ . However, since 1 percent tails are difficult to estimate accurately, standard deviations calculated in this manner may be unreliable; in addition, this method of computation makes no allowances for widely dispersed or sharply peaked distributions.* Hence, the resulting curve often grossly distorts the analyst's actual feelings. Furthermore, this restriction on the standard deviation seriously reduces the variety of shapes which are inherent in the beta family.

To overcome such shortcomings, it has been decided in the proposed technique to specify a fourth number, the 80 percent central range. The 80 percent central range, $C_R$ , represents the range of the estimated variable which contains 80 percent of the probability. It is calculated by subtracting the lower end of the range (the lower 10 percent tail) from the upper end of the range (the upper 10 percent tail). By permitting the analyst the freedom of specifying $C_R$ , more representative distributions often can

*Statisticians refer to this phenomenon as the distribution's kurtosis.

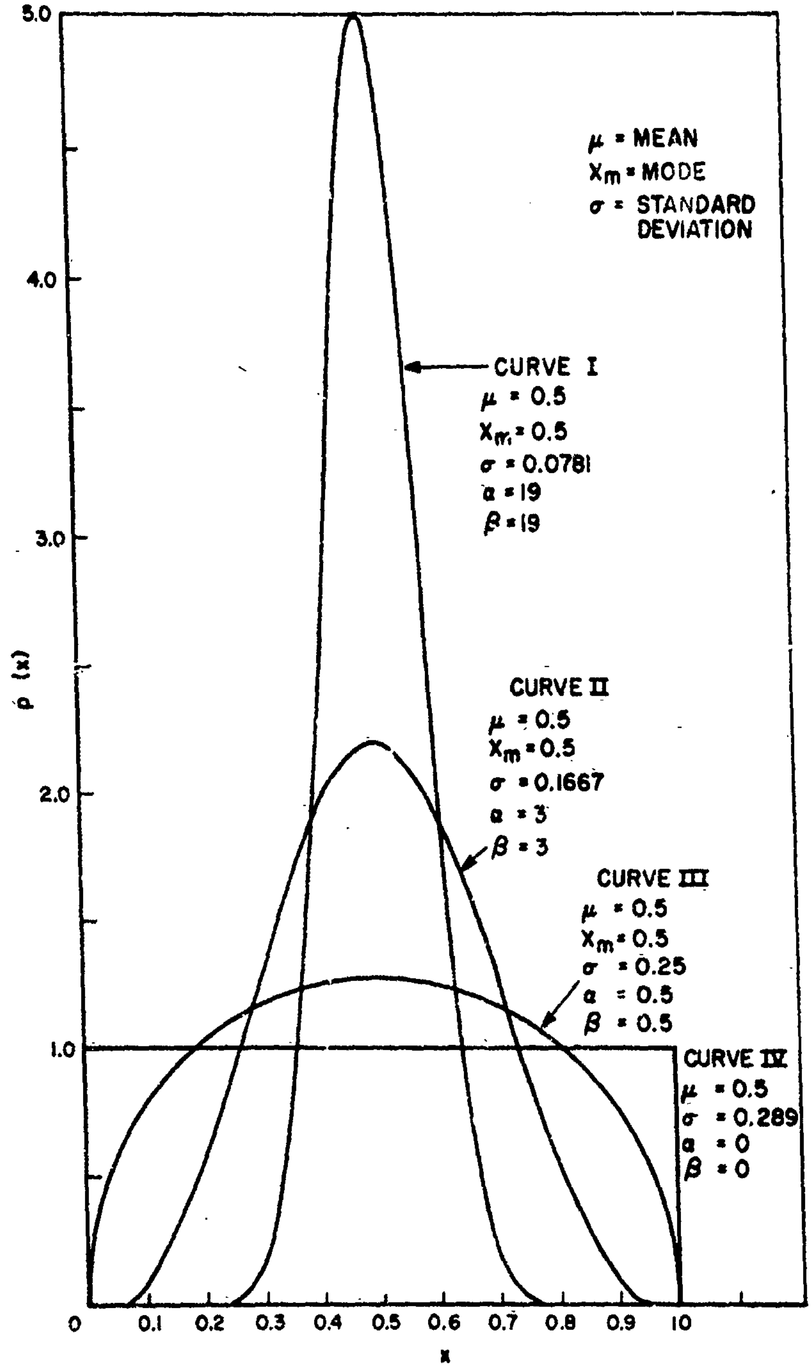



Figure 1. Symmetrical Beta Distributions

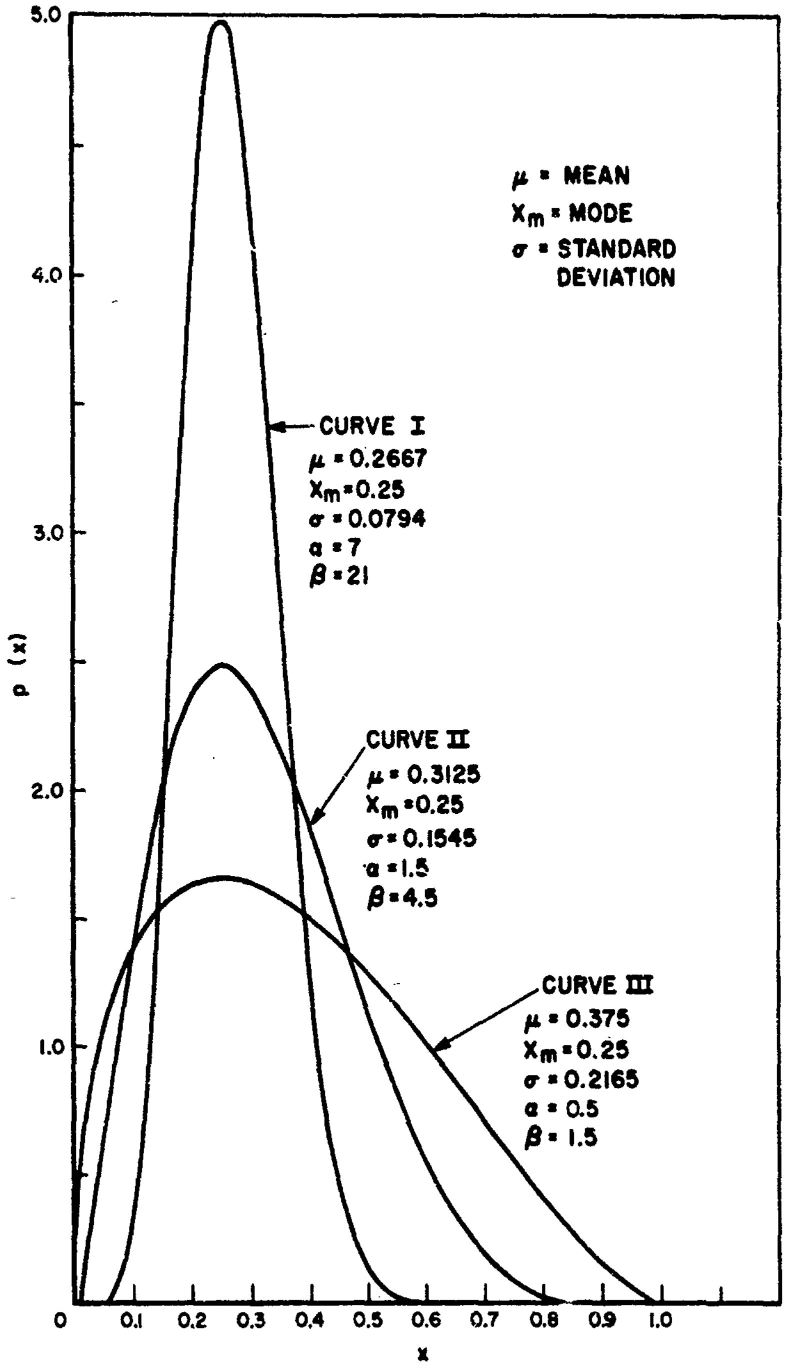



Figure 2. Skew Beta Distributions

result. For example, the rectangular, normal, and gamma distributions may all be accurately portrayed.* Figures 1 and 2 are provided to demonstrate some of the varied shapes that the beta distribution can assume.

Hence, for each cost element, the analyst provides four values, $X_L$ , $X_H$ , $X_P$ , and $C_R$ , to describe the uncertainty.** This information is based on the analyst's assessment of the probable sources and magnitude of the uncertainties. For such assessment, the analyst not only uses relevant, processed, historical data, but obtains and evaluates the opinions of experts and takes into account any discrepancies.

At this point it might be well to discuss the analyst's willingness to provide more data than in the past. It is quite probable that the analyst's reaction to such a situation will be related to his integrity. If he sincerely believes that extensive uncertainties enshroud the value of a particular cost element, it will be very difficult for him to specify a single value. Hence, it is the author's contention that when uncertainty exists, the conscientious cost analyst will appreciate the opportunity of being able to stipulate ranges.

## DETERMINING UNCERTAINTY IN TOTAL COSTS

Random variables representing cost elements may be combined in accordance with a cost structure to produce a random variable to represent the total system cost. By knowing (1) the distributions of the cost elements,

---

* The linearly scaled beta function is perhaps more generally known as a Pearson I curve. The gamma and normal are limiting forms and hence can be approximated with any desired accuracy. The rectangular distribution is a special case of the beta.

**When the uncertainty is insignificant, only one value need be specified. This value should be set equal to the mean and all higher moments can be taken equal to zero.

(2) the nature of any statistical dependence which may exist among them, and (3) the functional relationships which reflect the cost structure, the distribution for the total system cost may be determined.

To satisfactorily treat uncertainty in advanced system costs, it is considered adequate to represent the cost structure by products and sums of random variables where multiplied variables are independent and added variables are either independent or linearly dependent. The concept of linear dependence, as used in this paper, is defined in Appendix I.

The additive-multiplicative restriction on the cost structure is not believed to be a very serious handicap, since, in the author's experience, these operations are most frequently encountered in costing advanced electronic command and control systems. However, it is possible to process more general functional relationships, as described in Appendix II.

The limitations placed on intervariable dependence are also believed not to be highly restrictive. Appendix I shows that in many situations dependence can be avoided completely by the proper choice of cost element variables. However, in those instances where this expedient is not possible, linear dependence, in the author's opinion, will generally be adequate to reflect the statistical relationship among variables.

As demonstrated in Appendix I, the sum of linearly dependent variables may be expressed as sums and products of independent auxiliary random variables. The appendix further describes the information which the analyst may provide to determine these auxiliary variables. Thus, in a great many circumstances, the author believes that the cost structure may be satisfactorily represented by the sums and products of independent random variables.

It is well known that when independent random variables are added the distribution of their sum can be determined by convolving the component distributions. Similarly, when independent random variables are multiplied, the distribution of their product can be determined by an analogous process (which is portrayed in Appendix II). Nevertheless, since the input data about the uncertainty are only approximate, it was felt that such elaborate operations were not warranted. Instead, the decision was made to use moments to summarize the distributions. This procedure is discussed in the following section.

## UTILIZATION OF MOMENTS TO GENERATE THE DISTRIBUTION OF TOTAL COSTS

To treat uncertainty in PERT, the mean and standard deviation for each element distribution is computed and then used to compute the mean and standard deviation for the total. The central limit theorem, which states that under suitable conditions the distribution representing the sum of random variables tends to the normal, is then customarily invoked. Since the normal distribution is completely specified by the mean and standard deviation, the distribution of the total is thus determined.

The degree to which the normal faithfully portrays the distribution of the sum depends on many factors, including the number of variables added, the shapes and relative sizes of their distributions, and the degree of dependence among them. It is not clear, therefore, that total system costs will always be adequately represented by the normal, and it is even less apparent that the distribution of intermediate levels of cost aggregation will be adequately represented. The mean and standard deviations themselves provide no clue on the reliability of the normal approximation.

Figure 3 provides a graphic portrayal of the kinds of probability statements that can be made for a normal distribution. It also shows the upper bound for confidence statements that can be made for any distribution. In addition, it depicts an estimated upper bound for unimodal distributions.

Because confidence statements for the normal are so much tighter than for either of the indicated upper bounds, use of the latter bounds will frequently grossly overstate the uncertainty. Furthermore, the mean and standard deviation provide no insight about skewness. Thus, it is desirable to generate more information about the uncertainty in cost aggregations before attempting to construct a distribution.

To cope with the above situation, it was decided to use the first four, instead of just the first two, moments (such as the mean and standard deviation). In this regard, additive and multiplicative moment sets were employed. These are defined in the glossary and their properties are demonstrated in Appendix III. The $n^{th}$ additive moment for a sum of independent random variables is simply computed by adding the $n^{th}$ additive moments for each of the variables. Equivalently, the $n^{th}$ multiplicative moment for a product of independent random variables is the product of the $n^{th}$ multiplicative moments of each variable.

The distribution for total system cost (and for intermediate levels of cost aggregation) is constructed by determining the linearly scaled beta* (a four-parameter function) which has the same first four moments. The various mathematical operations that are involved in this process are

*If the beta distribution cannot satisfy these moments, the appropriate member function from the Pearson family may be used (see Table 43, Ref. [7]).

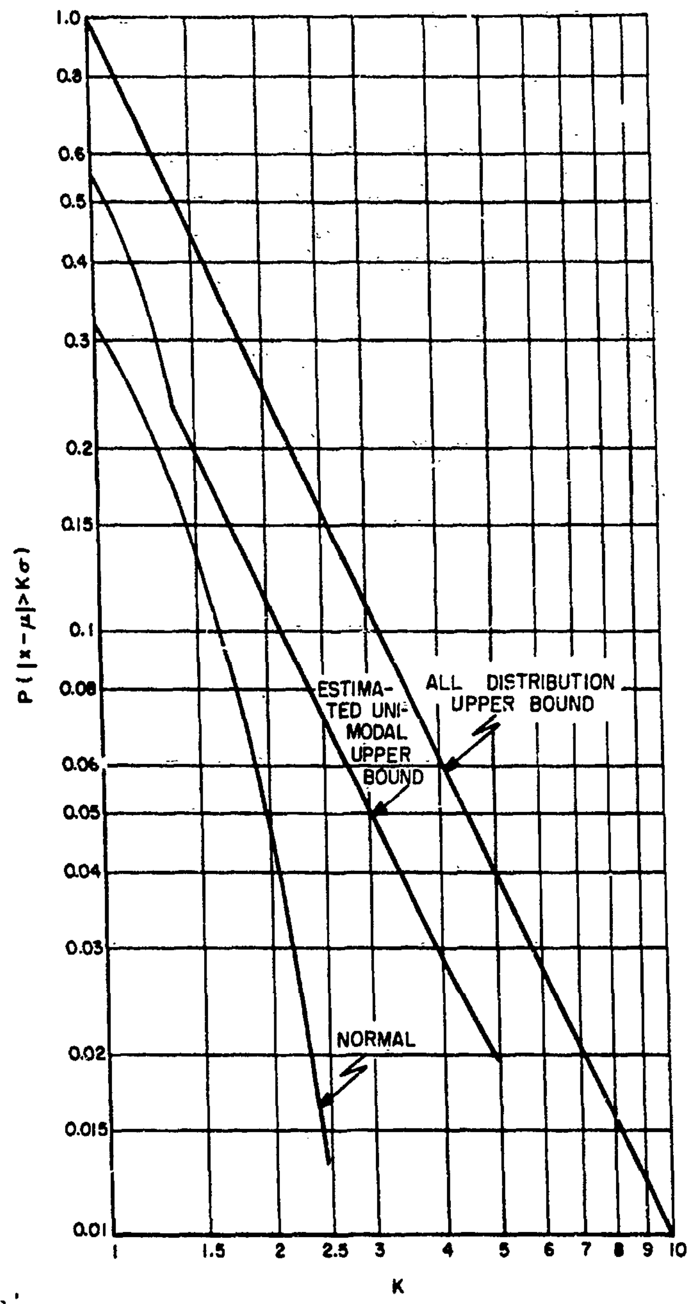



Figure 3. Probability that a Random Variable is More than Kσ Away from the Mean

described in Appendix IV. The integrity of this procedure is demonstrated in Appendix V by comparing the resultant distributions with the theoretical curves.

## SECTION III

## COMPUTER PROGRAM

### INTRODUCTION

A program has been written for the 7090 to facilitate the translation of the analyst's feelings about the uncertainty in each cost element into the parameters of an approximating linearly-scaled beta distribution for the cost total and sub-totals. These parameters are used to generate upper and lower bounds to a cumulative probability interval of any desired size. The program is composed of three discrete parts which are described below.

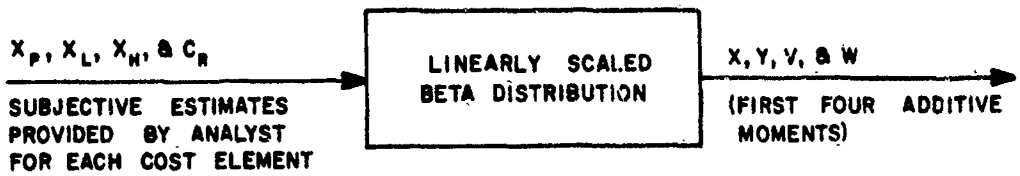
The reader is referred to Appendix VI for a more detailed description of the program, including an explanation of the mechanics involved to operate it on the MITRE 7090 Facility.

### PROGRAM I

Program I accepts as inputs those four parameters of the linearly scaled beta distribution which are provided by the analyst. The reader will recall that these are the most probable value, the lowest and highest values, and the 80 percent central range. With this information, the four additive moments are computed and presented as an output. The input is put on punch cards and the output is printed. The reader is referred to Figure 4 for a graphic description of this program.

### PROGRAM II

Program II accepts the additive moments computed by Program I for all of the cost elements and computes an equivalent set of moments for




$X_P$ MOST LIKELY VALUE (MODE)

$X_L$ & $X_H$ LOWER & UPPER 1% TAILS

$C_R$ 80% CENTRAL RANGE (UPPER 10% TAIL MINUS LOWER 10% TAIL)

X = 1st ORIGIN MOMENT = MEAN (EXPECTED VALUE)

Y = 2nd CENTRAL MOMENT = VARIANCE (STD DEVIATION$^2$)

V = 3rd CENTRAL MOMENT

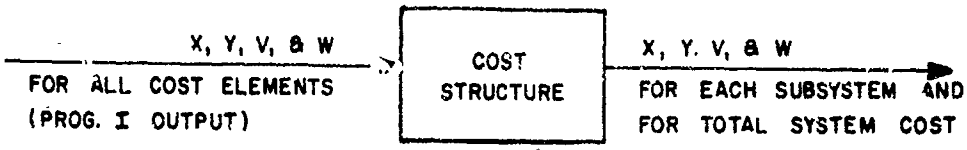
W = 4th CENTRAL MOMENT minus $Y^2$

Figure 4. Graphic Representation of Program I

cumulative costs at all desired levels of aggregation. An effort was made to keep the program sufficiently flexible so that a large variety of arithmetic arrangements in the cost structure could be handled. The input is inserted on punch cards. The output, specifying the additive moments of aggregated costs, is furnished on punch cards (which can be used as an input to Program III) and is also printed. Figure 5 depicts the program.

## PROGRAM III

Program III converts the additive moments computed by Program II into the parameters (see Glossary) for the linearly scaled beta distribution.




[RESTRICTION ON COST STRUCTURE:
SUMS AND PRODUCTS OF INDEPENDENT RANDOM VARIABLES]

Figure 5. Graphic Representation of Program II.

It further determines bounds for any $\gamma$ cumulative probability interval by finding limits for the $1 - \gamma/2$ tails.* Inputs are on punch cards and the output is printed. Figure 6 depicts this program.

*Exact upper and lower bounds to these tails can be computed directly from the first four moments to reveal the maximum error in the beta approximation. [8]

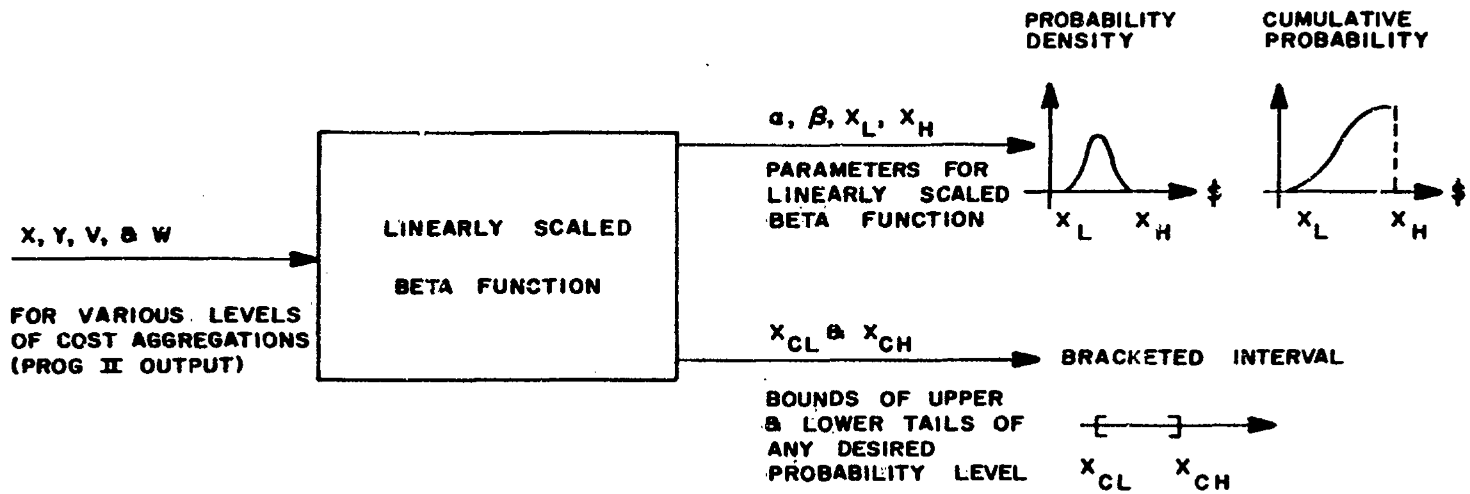



Figure 6. Graphic Representation of Program 111

## SECTION IV

## USEFULNESS OF THE RESULTS

Information provided by the technique proposed herein furnishes a better single value (as compared to conventional estimates) to represent the most likely cost of a military system and, in addition, furnishes a quantitative statement about the nature and extent of the uncertainty.

Conventional estimates typically introduce the most likely value for each cost element into the cost structure to determine the system's total economic implications. Nevertheless, since the distributions for the cost elements are generally skewed upwards, the single value thus computed for the system often has very little likelihood of occurring and may constitute a serious understatement of the true cost. When a single value is required from the cost analyst, the mean or the modal value (which are frequently relatively close for the total system cost) is generally a more meaningful[2] figure to specify. In systems which are characterized by considerable research and development, these latter values have been shown to exceed the conventional estimate by more than 30 percent.

Bracketed probability intervals provide a quantitative measure of the range of the uncertainty in aggregate costs. Such intervals provide insight on whether or not the accuracy of a cost estimate satisfies the requirements of its intended application. The decision maker may establish in advance the accuracy criterion which must be met. The bracketed interval would disclose whether the criterion is satisfied.

Bracketed probability intervals for subsystem costs and other intermediate levels of aggregation serve to identify areas of major cost uncertainty. These are areas which might benefit from subsequent efforts to refine costs. However, consideration should be given to the sensitivity of the uncertainties to more intensive study, since substantial additional probing will not materially reduce certain types of uncertainties.

The probability function can be used to make computations which may provide extremely useful information to the decision maker. For example, when two or more equal effectiveness, alternative configurations are being considered for acquisition, one can either immediately select the system with the lowest expected cost or defer a decision, pending the collection of more information. By knowing the entire probability function, it is possible to compute (1) the probability that the alternative having the lowest ultimate cost will not be selected by using the lowest mean criterion, and (2) the expected size of the cost premium which results from not having perfect information.[2] This expected loss (premium) can be used as an index of the maximum effort that should be devoted to refining cost or to pursuing exploratory R&D.[9] Conversely, this expected loss represents the maximum expected gain from deferring immediate action.

## SECTION V

## VALIDITY OF RESULTS

The reader is no doubt well aware of the fact that a conventional cost estimate is no better than the quality of its inputs. The analyst's ability to identify all cost contributing elements and to properly assess their economic implications bears directly on the validity of the end product. Similarly, in appraising the sources and magnitudes of the uncertainties which influence each element's cost, there is no satisfactory substitute for reliable data. The cost analyst must carefully and astutely search out all significant factors which give rise to system element cost variability and reflect the magnitude of their effect. The author wishes to point out, however, that although this does place an added burden on the analyst, he will probably derive greater satisfaction from pursuing this task than trying to generate a single number in the face of imperfect information.

It is true that the analyst's inputs to describe the uncertainty are subjective. Nevertheless, because of the general unavailability of completely relevant historical data from which the cost of advanced systems can be accurately gauged, single value estimates, themselves, contain varying degrees of subjectivity. What is perhaps more important, however, is the fact that although cost analysts in the past have been fully aware of the presence and character of element uncertainties, the means did not exist to properly assess their cumulative impact on the total system cost. Hence, more often than not, only an extremely unreliable conjecture about total cost uncertainty could be passed on to decision makers. By applying the proposed technique, the cost analyst is permitted to focus his attention on appraising

the uncertainties surrounding the individual cost elements, an area where he is believed more qualified to judge uncertainty; and standard statistical methods are then employed to measure their aggregate implications on the system total.

If the analyst inadvertently neglects to consider significant sources of uncertainty or has even forgotten to include some cost elements, the results, naturally, would understate the magnitude of the uncertainty and bias the distribution. Oversight of relevant cost areas is, of course, an error which has just as serious an effect on single-value estimates, and hence, much work has been and is being done to minimize this possibility. Similarly, work is needed in developing a capability to realistically appraise the sources and magnitude of uncertainties.

It is the author's contention that any information the analyst can provide regarding the magnitude of the uncertainty (even if occasionally in the form of an educated guess) is better than no knowledge at all. The proposed technique, in its present form, is regarded as merely a first step in the direction of providing decision makers with more meaningful information about the variability in total system costs.

## SECTION VI

## AREAS OF FUTURE WORK

Many refinements of the technique described herein will probably germinate from the experience gained in its successive application. Two specific areas where more intensive study seems to offer large potential rewards are: (1) development of techniques to aid in obtaining quantitative information about the cost uncertainty from experts; and (2) development of analytical tools that will permit fuller utilization of the generated statements about the uncertainty in making planning decisions.

## APPENDIX I

## DEPENDENCE AMONG VARIABLES

### AVOIDANCE OF DEPENDENCY

In a typical cost estimate, there are certain cost items which one normally would expect to be dependent. An example is mission equipment cost, initial spares cost, and maintenance cost. Nevertheless, if the total of these three items is determined as below

$$N \times U_c \left[1 + .01P_M \left(1 + \gamma_S\right)\right] ,$$

where

$N$ = number of units;

$U_c$ = equipment cost;

$P_M$ = percentage maintenance costs are of equipment costs; and

$\gamma_S$ = ratio of spare costs to maintenance costs,

the component variables in the above expression may frequently be regarded as independent. When the above technique can be employed, it is possible to avoid other means of specifying the nature of the dependency.

### LINEAR DEPENDENCE*

#### Definition and Properties

When two cost elements are dependent, the relationship between the elements can often be satisfactorily expressed by the following equality:

* The reader should not confuse this concept with the term as applied to sets of matrices or polynomials. Dependence here is concerned with the statistical relationships among random variables; see Glossary.

$$x_2 = \left(K_a + \varepsilon\right)\left(x_1 - \mu_1 + K_b\right) + K_c \quad ,$$

where

| | | |
|---|---|---|
| $x_1$ and $x_2$ | = | random variables representing the cost elements; |
| $\varepsilon$ | = | a random variable of zero mean and independent of $x_1$ ; |
| $\mu_1$ | = | the mean of $x_1$ ; and |
| $K_a$, $K_b$, and $K_c$ | = | constants. |

In the above situation, $x_1$ and $x_2$ are said to be (for the purposes of this report) linearly dependent. The conditional distribution for $x_2$ given $x_1$ , has the following properties: (1) when $x_1$ equals its mean, the distribution for $x_2$ can be specified without constraint; and (2) the distributions for other values of $x_1$ are represented by a linear shift and expansion (or contraction) of the coordinates.* The reader will note that the sum of linearly dependent random variables $x_1$ and $x_2$ becomes:

$$x_2 + x_1 = \left(K_a + 1 + \varepsilon\right)\left(x_1 + K_b - \mu_1\right) - K_b + K_c + \mu_1 \quad ,$$

which is composed of independent random variables.

Determination of Independent Variables

The variables $K_a$ , $K_b$ , $K_c$ and $\varepsilon$ are determined on the basis of information furnished by the analyst. Although there is more than one manner

*Thus

$$P\left[x_2/\left(x_1 = \mu_1\right)\right] = P\left(\frac{x_2 - a}{b+1} \Big/ x_1\right)$$

where

$$a = K_a\left(x_1 - \mu_1\right); \; b = \frac{1}{K_b}\left(x_1 - \mu_1\right)$$

of eliciting the necessary information, only one will be considered in this report.

The analyst is called on to describe the conditional distribution for $x_2$, given that $x_1$ is at its mean value, $\mu_1$. He can do this by specifying values for the parameters of the linearly scaled four-parameter beta. From these parameters, the four additive moments $A_1$, $A_2$, $A_3$ and $A_4$ can be determined. The analyst is further requested to specify the amount by which the conditional mean of $x_2$ will change when $x_1$ deviates from $\mu_1$ by some fixed amount. The ratio of these two items (change in $\mu_{2/1}$/amount $x_1$ deviates) will be called $m_\mu$. Finally, the analyst will be asked to state the amount by which the conditional standard deviation of $x_2$ will change when $x_1$ deviates from $\mu_1$. The ratio of these items (change in $\sigma_{2/1}$/ amount $x_1$ deviates) will be called $m_\sigma$.

The resulting relationships become:

$$K_a = m_\mu \qquad K_b = \frac{\sqrt{A_2}}{m_\sigma} \qquad K_c = A_1 - K_b K_a$$

$$E_1 = 0 \qquad E_2 = m_\sigma^2 \qquad E_3 = \frac{A_3}{K_b^3} \qquad E_4 = \frac{A_4}{K_b^4} ,$$

where the $E_k$ terms are the additive moments of $\varepsilon$.

In the special case where $m_\sigma$ is believed to be zero, the following relation should be used:

$$x_2 = \mu_2 + K_a\left(x_1 - \mu_1\right) + \varepsilon .$$

In this case,

$$\mu_2 = A_1 \quad K_a = m_\mu \quad E_1 = 0 \quad E_2 = A_2 \quad E_3 = A_3 \quad E_4 = A_4 .$$

If the analyst prefers to specify $m_\chi$ (the ratio of the change in the conditional mode of $x_2$ to the deviation of $x_1$ from $\mu_1$) to $m_\mu$, the following equation may be used:

$$m_\mu = m_\chi + \Delta \cdot \frac{m_\sigma}{\sigma} ,$$

where

$\Delta$ is the difference between the conditional mean and mode of $x_2$, given $x_1 = \mu_1$.

## More General Form of Linear Dependence

Let $x_1$, $x_2$, ..... $x_n$ be random variables which are linearly dependent on the independent random variables $a_1$, $a_2$, ..... $a_m$. It is desired to express the sum

$$\sum_{k=1}^{n} x_k,$$

as the sum and product of independent random variables. If each $x$ can be expressed as

$$x_k = \sum_{j=1}^{m} \left(A_{kj} + \epsilon_{kj}\right)\left(a_j - \mu_j + B_j\right) + C_{kj} ,$$

where

$A_{kj}$, $B_j$, and $C_{kj}$ = constants;

$u_j$ = the mean of $a_j$; and

$\epsilon_{kj}$ = a random variable of zero mean and independent of all a's and other $\epsilon$'s;

then the sum may be expressed by the following:

$$\sum_{k=1}^{n} x_k = \sum_{j=1}^{m} \left\{ \left(a_j - \mu_j + B_j\right) \left[ \sum_{k=1}^{n} \left(A_{kj} + \varepsilon_{kj}\right) \right] + \sum_{k=1}^{n} C_{kj} \right\} ,$$

which is the desired result.

Dependence Among Additive-Multiplicative Combinations of Independent Elements

The situation sometimes arises in which independent random variables reappear in more than one expression. A simple example is as follows:

$$x_T = x_1 x_2 + x_1 x_3 + x_1 x_4 ,$$

where the $x_k$ terms are independent random variables.

The terms in the above example, therefore, are not independent. This situation, however, may easily be handled by factoring $x_1$ so that

$$x_T = x_1 \quad x_2 + x_3 + x_4 .$$

More complex situations, however, such as

$$x_T = x_1 x_2 + x_2 x_3 + x_1 x_3 ,$$

cannot be dealt with as simply since complete factorization[10] is not possible. In this case it is necessary to take advantage of the fact that the expected value for the sum of random variables, whether or not they are dependent, may be computed by adding their expected values. The reader will recall that the $k^{th}$ multiplicative moment for $x_T$ is

$$M_k \;\; x_T \;\; = \alpha^{(k)} \;\; x_T \;\; = E \;\; x_T^k ,$$

where E designates the expected value.

Hence

$$M_k\left(x_T\right) = E\left(x_1x_2 + x_2x_3 + x_1x_3\right)^k \quad .$$

The preceding expression may be evaluated by expanding the argument and then substituting for the origin moments involved. For example,

$$\begin{aligned}
M_2\, x_T &= E\left(x_1^2 x_2^2 + x_2^2 x_3^2 + x_1^2 x_3^2 + 2x_1^2 x_2x_3 + 2x_2^2 x_1x_3\right.\\
&\qquad \left. + 2x_3^2 x_1x_2\right)\\
&= M_2\left(x_1\right)M_2\left(x_2\right) + M_2\left(x_2\right)M_2\left(x_3\right) + M_2\left(x_1\right)M_2\left(x_3\right)\\
&\qquad + 2M_2\left(x_1\right) M_1\left(x_2\right) M_1\left(x_3\right)\\
&\qquad + 2M_2\left(x_2\right) M_1\left(x_1\right) M_1\left(x_3\right)\\
&\qquad + 2M_2\left(x_3\right) M_1\left(x_1\right) M_1\left(x_2\right)
\end{aligned}$$

This process is most readily executed by automated means.

## APPENDIX II

### PRODUCT OF INDEPENDENT RANDOM VARIABLES

If $x_1$ and $x_2$ are two random variables with a joint probability density function of $P_{12}\left(x_1, x_2\right)$ and z represents some function* f $x_1, x_2$, of these two variables, the probability density for z is

$$p(z) = \int_{-\infty}^{+\infty} P_{12}\left[g_1\left(z, x_2\right), x_2\right] \left| \frac{\partial x_1}{\partial z} \right| dx_2$$

where

$g_1\left(z, x_2\right) =$ the function that results from solving $f\left(x_1, x_2\right)$ for $x_1$ in terms of $z$ and $x_2$ ; and

$\left(\partial x_1/\partial_z\right) =$ the derivative of $g_1\left(z, x_2\right)$ holding $x_2$ constant.

In the case under consideration, $f\left(x_1, x_2\right) \equiv x_1 \cdot x_2$, $P_{12}\left(x_1, x_2\right) \equiv p_1\left(x_1\right) p_2\left(x_2\right)$, and $g_1\left(z, x_2\right) = z/x_2$ . Thus

$$p(z) = \int_{-\infty}^{+\infty} p_1\left(z/x_2\right)p_2\left(x_2\right)x_2^{-1}dx_2 \ .$$

If $p_1\left(x_1\right) \equiv p_2\left(x_2\right) \equiv 0$ when their arguments are either less than some non-negative number, a, or greater than a + 1, the distribution for z becomes

* This function must be differentiable and single-valued over the ranges of $x_1$ and $x_2$ where $p_{12}\left(x_1, x_2\right)$ is nonzero. If the inverse function $g_1\left(z, x_2\right)$ is not single-valued, then the integral for $p(z)$ must be appropriately broken into segments and the results added.

$$p_m(z) = \int_L^U x_2^{-1}\, p_1\left(z/x_2\right)\, p_2\left(x_2\right)\, dx_2 \quad ,$$

where

$L = a, \quad U = z/a \quad$ when $z \leq a\,(a+1)$ ; and

$L = \frac{z}{a+1}, \quad U = a+1, \quad$ when $z \geq a\,(a+1)$ .

and $a^2 \leq z \leq (a+1)^2$.

If we now define the variables and functions as follows:

$$\ln x_1 = u_1 \ln \frac{a+1}{a} + \ln a, \;\; \ln x_2 = u_2 \ln \frac{a+1}{a} + \ln a, \;\; \ln z = w \ln \frac{a+1}{a} + 2 \ln a$$

$$p_1\left(x_1\right) = g_1\left(u_1\right),\; p_2\left(x_2\right) = g_2\left(u_2\right),\; p_m(z) = g_m(w) \quad ,$$

it follows that

$$g_m(w) = \ln\left(\frac{a+1}{a}\right) \int_{L_u}^{U_u} g_1\left(w-u_2\right)\, g_2\left(u_2\right)\, du_2 \quad ,$$

where

$L_u = 0, \quad U_u = w \quad$ when $w \leq 1$ ; and

$L_u = w-1, \quad U_u = 1 \quad$ when $w \geq 1$ .

The ranges of the variables are: $0 \leq u_1 \leq 1, 0 \leq u_2 \leq 1$, and $0 \leq w \leq 2$.

If $u_1$ and $u_2$ are considered as independent random variables, then $w = u_1 + u_2$. This relationship was utilized in Section IV when

computing the exact solution for those illustrative examples that involve products.

It becomes evident from the central limit theorem and the latter relationship that the limiting form as $n \to \infty$ of the distribution for

$$\prod_{k=1}^{n} x_k$$

(where the $x_k$ terms are independent random variables, with a common distribution defined for positive values) is log normal. The log normal is, of course, a skew unimodal distribution.

## APPENDIX III

### MOMENTS FOR SUMS AND PRODUCTS OF INDEPENDENT VARIABLES

The properties ascribed to the additive and multiplicative moments defined in the glossary will be demonstrated below.

### ADDITIVE MOMENTS

The additive moments $A_1$, $A_2$, $A_3$, and $A_4$ for the sum of independent random variables $x_1$ and $x_2$ are defined as follows:

$$A_1 = \alpha_s^{(1)} = E\left(x_1 + x_2\right) \quad ;$$

$$A_2 = \mu_s^{(2)} = E\left(x_1 - \mu_1 + x_2 - \mu_2\right)^2 \quad ;$$

$$A_3 = \mu_s^{(3)} = E\left(x_1 - \mu_1 + x_2 - \mu_2\right)^3 \quad ; \text{ and}$$

$$A_4 = \mu_s^{(4)} - 3\left[\mu_s^{(2)}\right]^2 = E\left(x_1-\mu_1+x_2-\mu_2\right)^4 - 3E^2\left(x_1-\mu_1+x_2-\mu_2\right)^2 \quad .$$

where

$\alpha_s^{(k)}$ = the $k^{th}$ origin moment of the sum

$\mu_s^{(k)}$ = the $k^{th}$ central moment of the sum

$\mu_1$ and $\mu_2$ = the means of $x_1$ and $x_2$, respectively

$E$ = the expected value

Evaluation of the right-hand members of the above equalities yields

$$E\left(x_1+x_2\right) = E\left(x_1\right)+E\left(x_2\right)=\alpha_1^{(1)}+\alpha_2^{(2)} \quad ;$$

$$E\left(x_1-\mu_1+x_2-\mu_2\right)^2 = E\left(x_1-\mu_1\right)^2+2E\left(x_1-\mu_1\right)E\left(x_2-\mu_2\right)+E\left(x_2-\mu_2\right)^2 ,$$

$$= E\left(x_1-\mu_1\right)^2+E\left(x_2-\mu_2\right)^2 ,$$

$$= \mu_1^{(2)}+\mu_2^{(2)} \quad ;$$

$$E\left(x_1-\mu_1+x_2-\mu_2\right)^3 = E\left(x_1-\mu_1\right)^3+3E\left(x_1-\mu_1\right)^2E\left(x_2-\mu_2\right)$$

$$+3E\left(x_1-\mu_1\right)E\left(x_1-\mu_1\right)^2+E\left(x_1-\mu_1\right)^3,$$

$$= E\left(x_1-\mu_1\right)^3+E\left(x_1-\mu_1\right)^3 ,$$

$$= \mu_1^{(3)}+\mu_2^{(3)} \quad ;$$

$$E\left(x_1-\mu_1+x_2-\mu_2\right)^4-3E^2\left(x_1-\mu_1+x_2-\mu_2\right)^2=E\left(x_1-\mu_1\right)^4$$

$$+4E\left(x_1-\mu_1\right)^3E\left(x_2-\mu_2\right) ,$$

$$+6E\left(x_1-\mu_1\right)^2E\left(x_2-\mu_2\right)^2 .$$

$$+4E\left(x_1-\mu_1\right)E\left(x_2-\mu_2\right)^3$$

$$+E\left(x_2-\mu_2\right)^4$$

$$- 3\left[\mu_1^{(2)}\right]^2 - 6\mu_1^{(2)}\mu_2^{(2)} - 3\left[\mu_2^{(2)}\right]^2,$$

$$= \mu_1^{(4)} - 3\left[\mu_1^{(2)}\right]^2 + \mu_2^{(4)} - 3\left[\mu_1^{(2)}\right]^2 ,$$

where

$\alpha_1^{(k)}, \alpha_2^{(k)}$ = the $k^{th}$ origin moments of $x_1$ and $x_2$, respectively; and

$\mu_1^{(k)}, \mu_2^{(k)}$ = the $k^{th}$ central moments of $x_1$ and $x_2$ respectively.

It can be seen from the above, that the $k^{th}$ additive moment for a sum of independent random variables equals the sum of the $k^{th}$ additive moments of the variables.

MULTIPLICATIVE MOMENTS

The multiplicative moments $M_1$, $M_2$ $M_3$, and $M_4$ for the product of independent random variables $x_1$ and $x_2$ are defined as follows:

$$M_k = \alpha_p^{(k)} = E\left(x_1 \cdot x_2\right)^k ,$$

where

$\alpha_p^{(k)}$ is the $k^{th}$ origin moment of the product.

Evaluation of the right-hand member of the above equation yields:

$$E\left(x_1 \cdot x_2\right)^k = E\left(x_1\right)^k \; E\left(x_2\right)^k = \alpha_1^{(k)} \alpha_2^{(k)} ,$$

where

$\alpha_1^{(k)}$ and $\alpha_2^{(k)}$ are the $k^{th}$ origin moments of $x_1$ and $x_2$, respectively.

It is obvious, therefore, that the $k^{th}$ multiplicative moment for a product of independent random variables equals the product of the $k^{th}$ multiplicative moments of the variables.

CONVERSION BETWEEN MOMENT SETS

The following formulas show additive moments may be generated from multiplicative moments, and vice versa:

$$M_1 = A_1, \; M_2 = A_2 + A_1^2, \; M_3 = A_3 + 3A_1A_2 + A_1^3,$$

$$M_4 = A_4 + 3A_2^2 + 4A_1A_3 + 6A_1^2A_2 + A_1^4 \;;$$

$$A_1 = M_1, \; A_2 = M_2 - M_1^2, \; A_3 = M_3 - 3M_1M_2 + 2M_1^3$$

$$A_4 = M_4 - 3A_2^2 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4 \; .$$

## APPENDIX IV

## DETERMINATION OF THE PARAMETERS FOR THE APPROXIMATING LINEARLY SCALED BETA

Given the first four additive moments, a linearly scaled Beta distribution is uniquely defined. To determine the four parameters, $\alpha$, $\beta$, a, and b (see Glossary), the following procedure may be used.

1) Solve the following two equations simultaneously for $\alpha$ and $\beta$:

$$\frac{A_3^2}{A_2^3} = \frac{4(\alpha-\beta)^2\,(\alpha+\beta+3)}{(\alpha+1)\,(\beta+1)\,(\alpha+\beta+4)^2}$$

$$\frac{A_4}{A_2^2} + 3 = \frac{3(\alpha+\beta+3)\,\left[2\,(\alpha+\beta+2)^2 + (\alpha+1)\,(\beta+1)\,(\alpha+\beta-4)\right]}{(\alpha+\beta+3)\,(\beta+1)\,(\alpha+\beta+4)\,(\alpha+\beta+5)}$$

where

$\alpha < \beta$ when $A_3 > 0$;
$\alpha = \beta$ when $A_3 = 0$; and
$\alpha > \beta$ when $A_3 < 0$.

2) Substitute $\alpha$ and $\beta$ into the equations shown below:

$$a = A_1 - \sqrt{A_2}\,\frac{\mu_n}{\sigma_n} \qquad b = \frac{\sqrt{A_2}}{\sigma_n}$$

where

$$\mu_n = \frac{\alpha+1}{\alpha+\beta+2} \text{ , and}$$

$$\sigma_n = \frac{\sqrt{\frac{(\alpha+1)\ (\beta+1)}{\alpha+\beta+3}}}{\alpha+\beta+2} \ .$$

## APPENDIX V

## ILLUSTRATIVE EXAMPLES

### INTRODUCTION

The examples shown in this appendix demonstrate, under a variety of situations, the integrity of the approximating procedures (see Section II) built into the computer program. The results of the procedure are compared with the results obtained from actual performance of the convolution operation for the sum of independent random variables and with the operation described in Appendix II for the product of independent random variables. The latter operations were performed on an analog computer.

### EXAMPLE ONE

Four independent random variables describing the uncertainty in cost elements all have the same distribution. This distribution can be represented by a linearly scaled beta with the following parameters: the most likely cost is 100 thousand dollars, the most optimistic cost is 50 thousand dollars, the most pessimistic cost is 250 thousand dollars, and the central range is 110 thousand dollars. The uncertainty for the total of these four costs can be examined. A plot of the probability distribution for the total cost is presented in Figure 7.

### EXAMPLE TWO

Two independent random variables, describing the uncertainty in cost elements, have the same distribution, which can be represented by a linearly

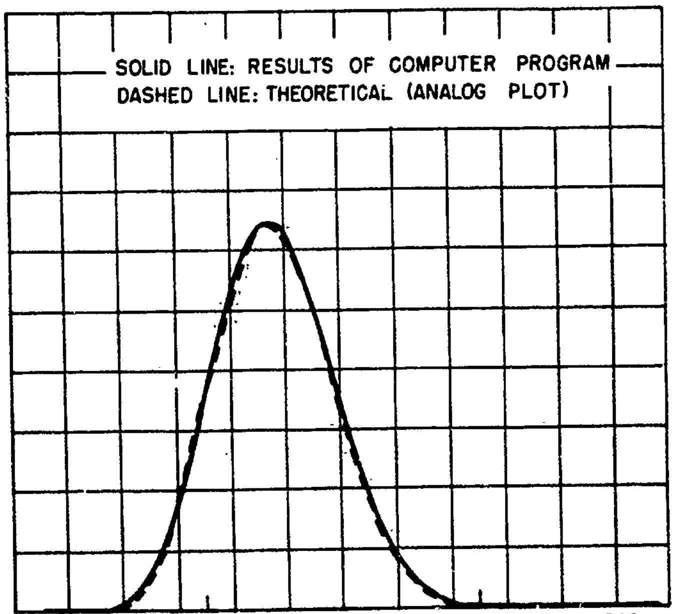



Figure 7. Plot of the Sum of Four Independent Random Variables with Beta Parameters

scaled beta. The probability distribution for the sum of the two costs is desired.

Four sets of parameters for the linearly scaled beta are considered. (The values for each parameter are shown in thousands of dollars.)

| Optimistic ($X_L$) | Most Probable ($X_P$) | Pessimistic ($X_H$) | Central Range ($C_R$) |
|---|---|---|---|
| 50 | 100* | 150 | 64 |
| 50 | 100** | 150 | 19 |
| 50 | 100*** | 250 | 110 |
| 50 | 100**** | 250 | 38 |

The distributions for total cost appear in Figures 8 to 11.

EXAMPLE THREE

Two independent random variables, describing the uncertainty in planning factors, have the same distribution which can be represented by a linearly scaled beta. The most pessimistic value for each of the random variables is 200 percent of the most optimistic value. The probability distribution for the product of the two costs is desired.

*The normalized form of this distribution appears as Curve III in Figure 1.

**The normalized form of this distribution appears as Curve I in Figure 1.

***The normalized form of this distribution appears as Curve III in Figure 2.

****The normalized form of this distribution appears as Curve I in Figure 2.

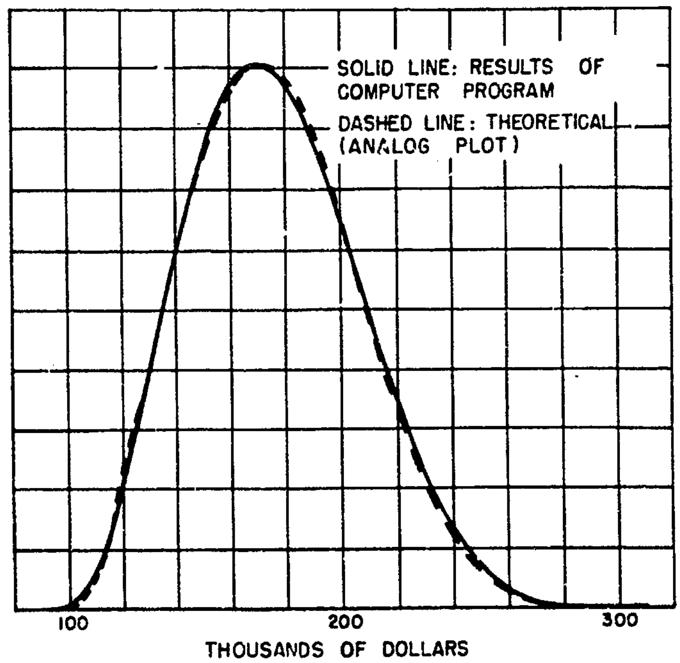



Figure 8. Plot of the Sum of Two Independent Random Variables with Beta Parameters

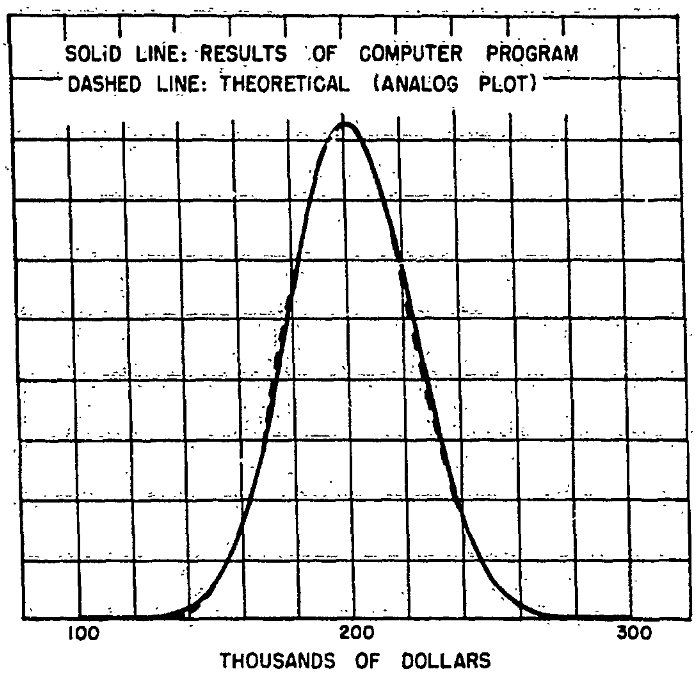



Figure 9. Plot of the Sum of Two Independent Random Variables with Beta Parameters

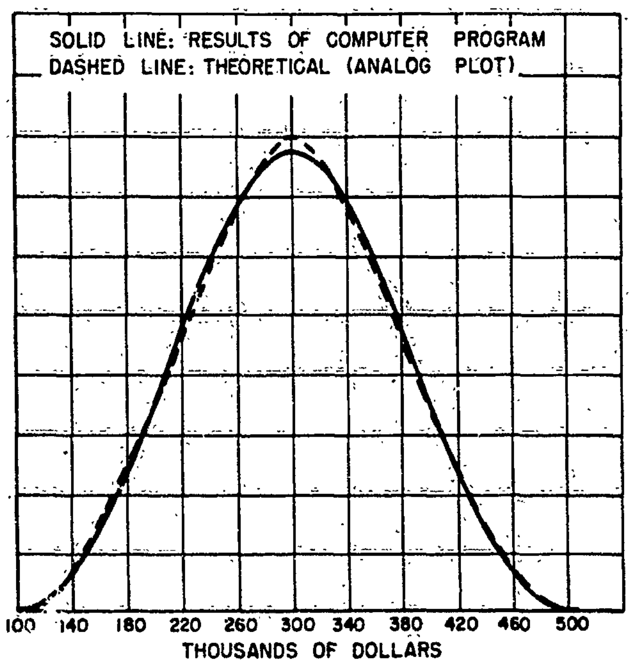



Figure 10. Plot of the Sum of Two Independent Random Variables with Beta Parameters

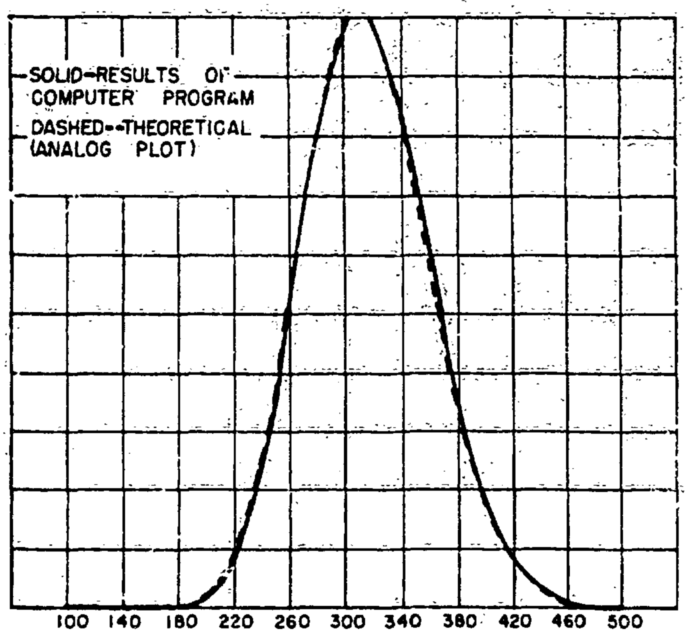



Figure 11. Plot of the Sum of Two Independent Random Variables with Beta Parameters

Four sets of parameters for the linearly scaled beta are considered.

| Optimistic ($X_L$) | Most Probable ($X_P$) | Pessimistic ($X_H$) | Central Range ($C_R$) |
|---|---|---|---|
| 1 | 1.5* | 2 | .64 |
| 1 | 1.5** | 2 | .19 |
| 1 | 1.25*** | 2 | .55 |
| 1 | 1.25**** | 2 | .19 |

The distributions for the products appear in Figures 12 to 15.

EXAMPLE FOUR

This example is similar to Example Three with the exception that the most pessimistic value in this case is 120 percent of the most optimistic figure. Four sets of parameters for the linearly scaled beta are considered:

| Optimistic ($X_L$) | Most Probable ($X_P$) | Pessimistic ($X_H$) | Central Range ($C_R$) |
|---|---|---|---|
| 5 | 5.5* | 6 | .64 |
| 5 | 5.5** | 6 | .19 |
| 5 | 5.25*** | 6 | .55 |
| 5 | 5.25**** | 6 | .19 |

The distributions for the products appear in Figures 16 to 19.

---

*Normalized form, Curve III in Figure 1.
**Normalized form, Curve I in Figure 1.
***Normalized form, Curve III in Figure 2.
****Normalized form, Curve I in Figure 2.

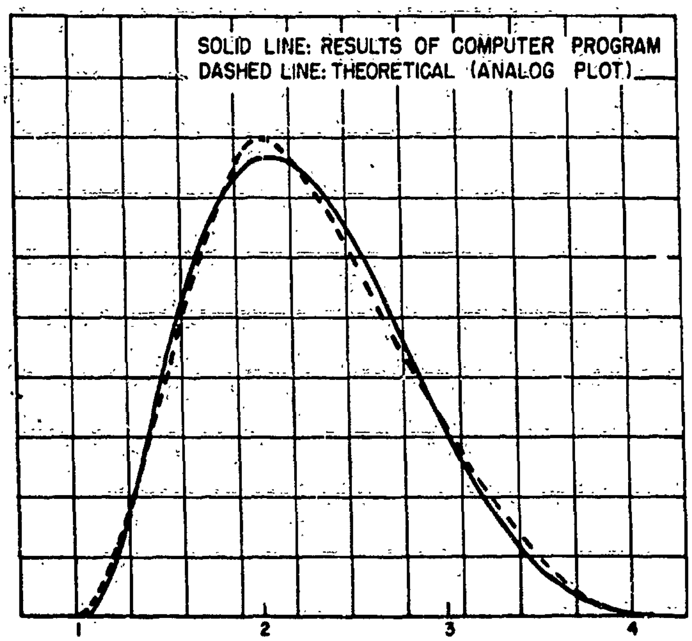



Figure 12. Plot of the Product of Two Independent Random Variables with Beta Parameters

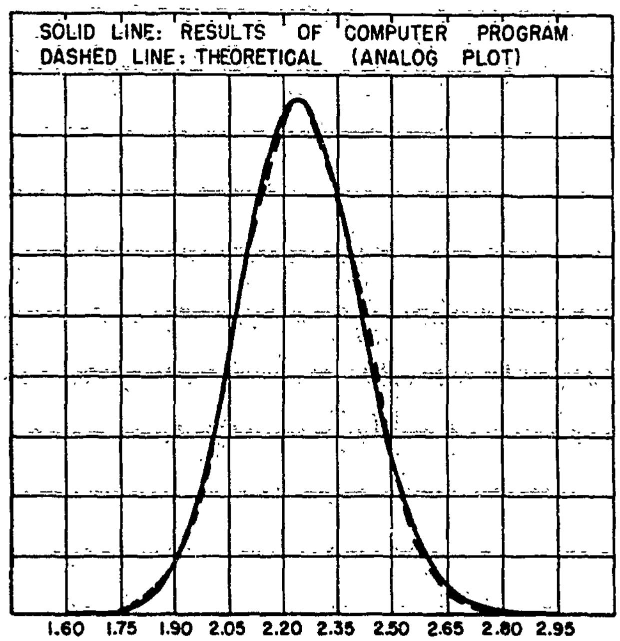



Figure 13. Plot of the Product of Two Independent Random Variables with Beta Parameters

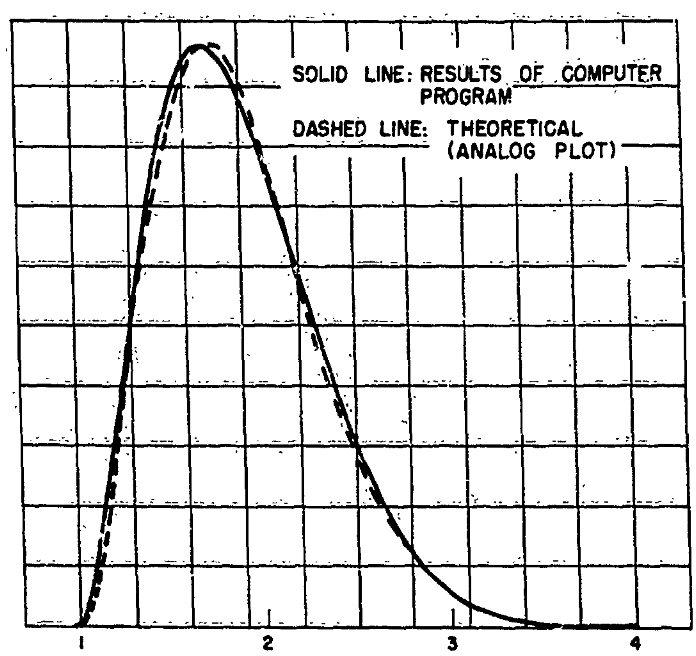



Figure 14. Plot of the Product of Two Independent Random Variables with Beta Parameters

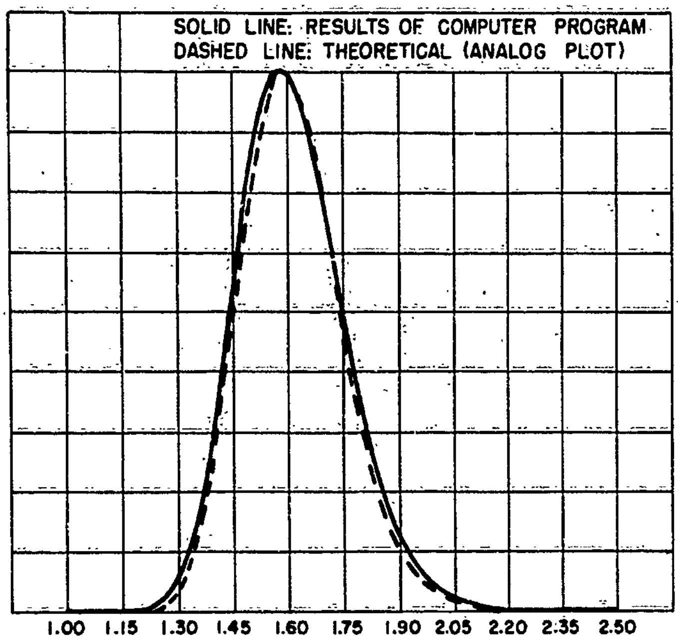



Figure 15. Plot of the Product of Two Independent Random Variables with Beta Parameters

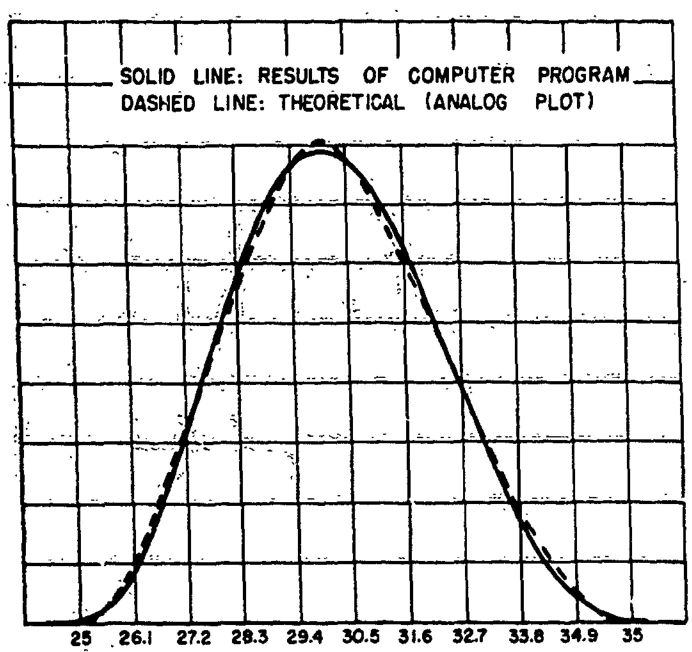



Figure 16. Plot of the Product of Two Independent Random Variables with Beta Parameters

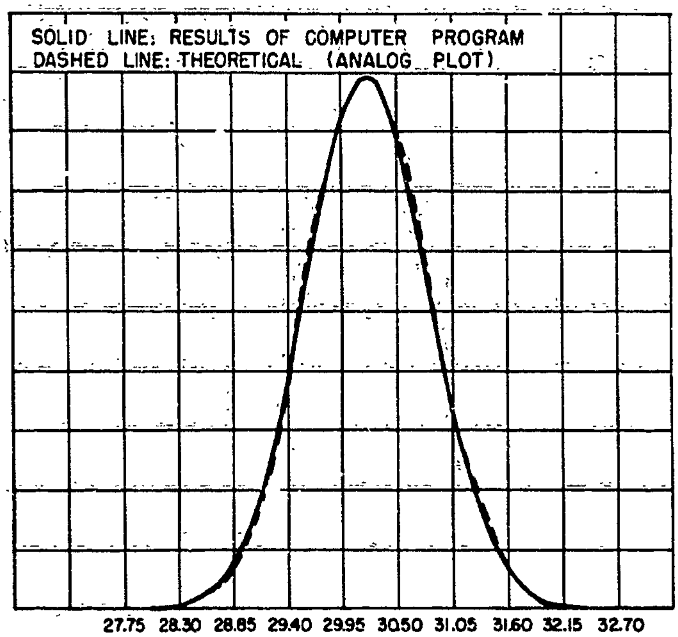



Figure 17. Plot of the Product of Two Independent Random Variables with Beta Parameters

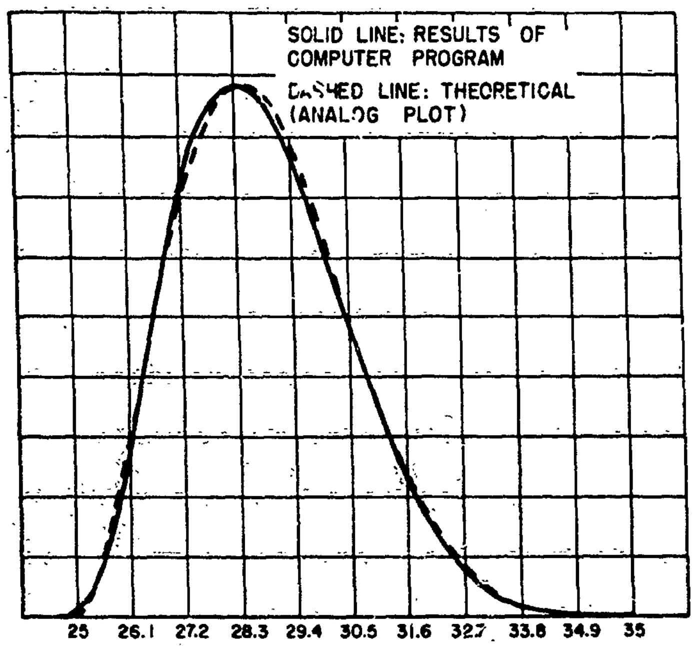



**Figure 18. Plot of the Product of Two Independent Random Variables with Beta Parameters**

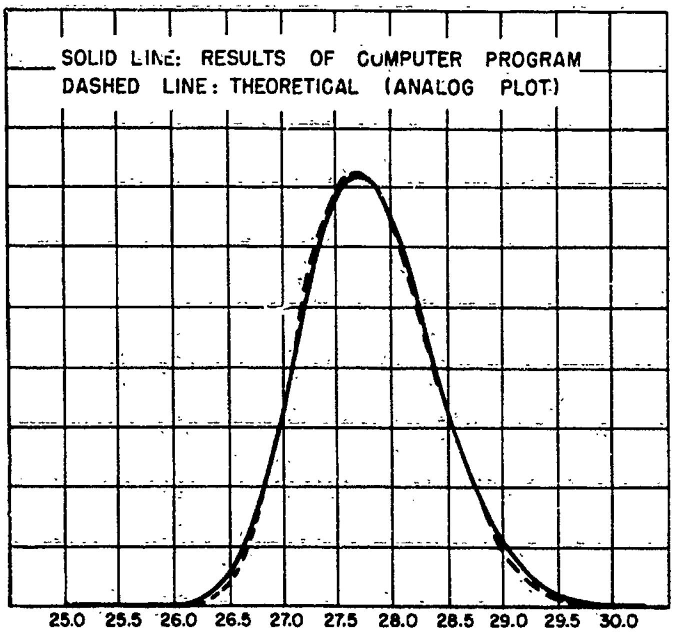



Figure 19. Plot of the Product of Two Independent Random Variables with Beta Parameters

## APPENDIX VI

## DESCRIPTION OF THREE 7090 FORTRAN PROGRAMS

### INTRODUCTION

This appendix describes a series of three 7090 FORTRAN programs (see Figure 20) which convert a cost analyst's subjective feelings about uncertainty in cost estimates for system elements into quantitative data to express the uncertainty in the total system cost.

### DESCRIPTION OF THE PROGRAMS

#### Program I

This program (see Figure 21) accepts as inputs the four parameters specified by the analyst for each element, and from these computes a linearly scaled beta distribution which approximates the distribution described by the parameters for the cost of the element. This distribution is presented in terms of the parameters describing the shape of the curve and the first four additive moments of the distribution.

The four input parameters are:

(a) $X_L$ - the lowest cost (lower 1 percent tail);

(b) $X_H$ - the highest cost (upper 1 percent tail);

(c) $X_P$ - the most probable cost (mode); and

(d) $C_R$ - the length of the 80 percent central range (upper 10 percent tail minus lower 10 percent tail).

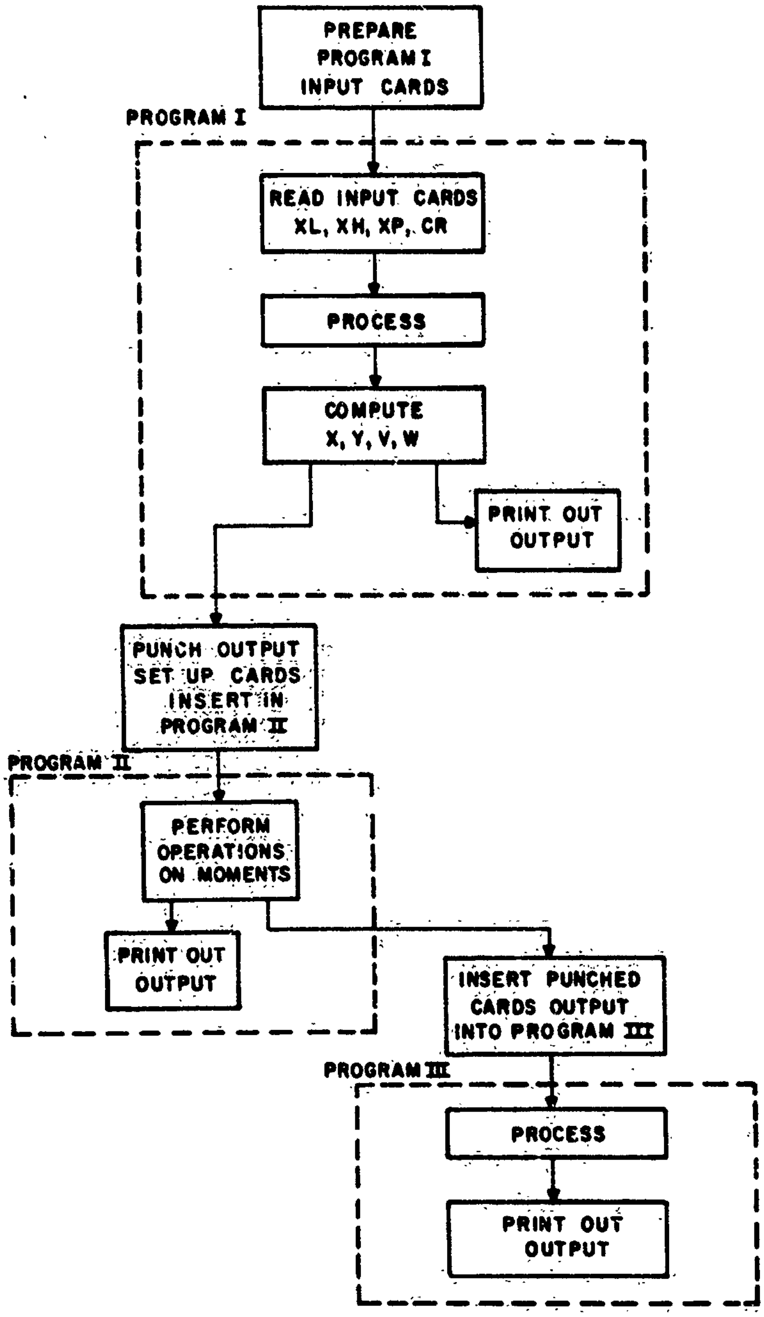



Figure 20. Block Diagram for Programs I, II and III

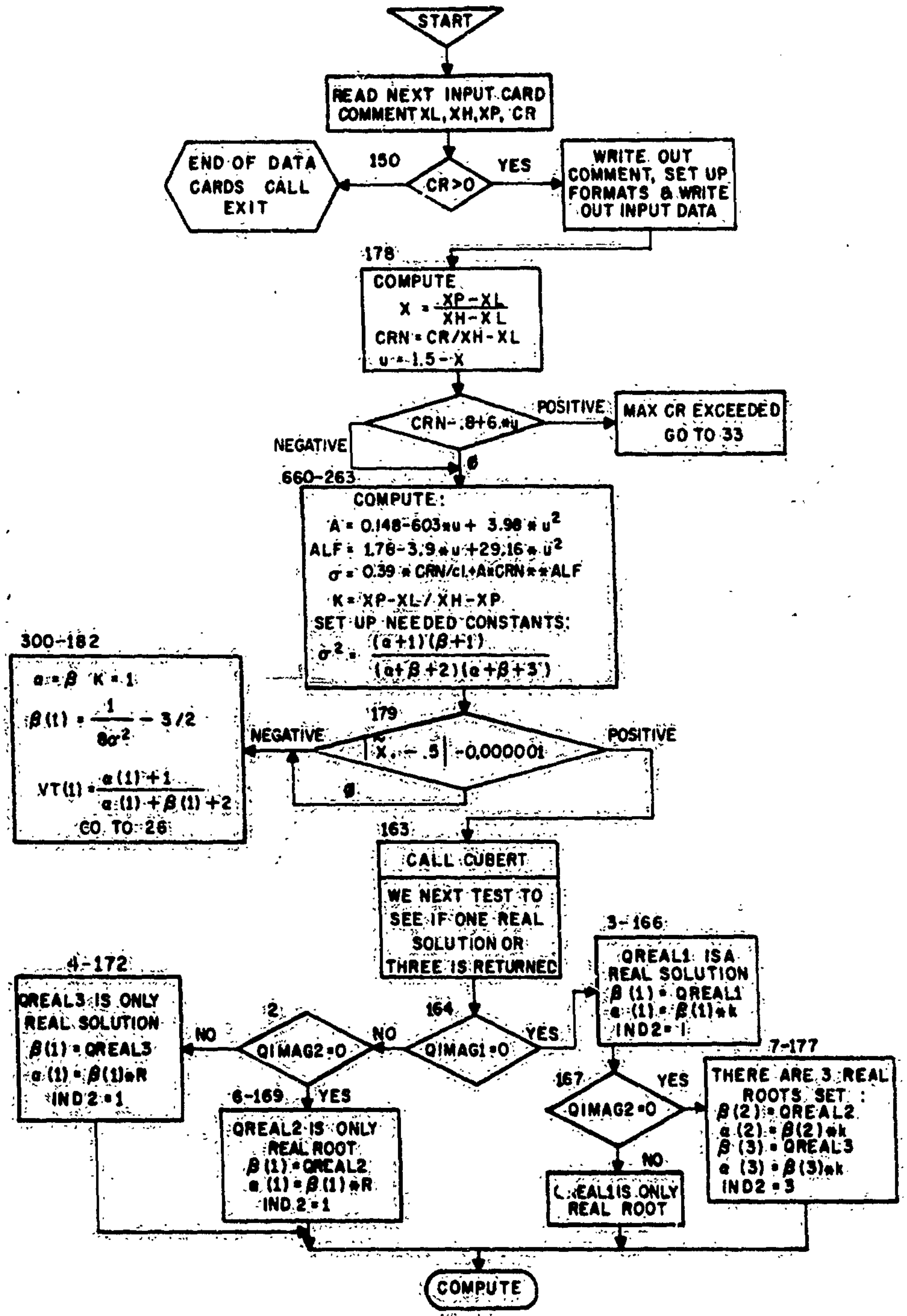



Figure 21. Program 1: Main Routine

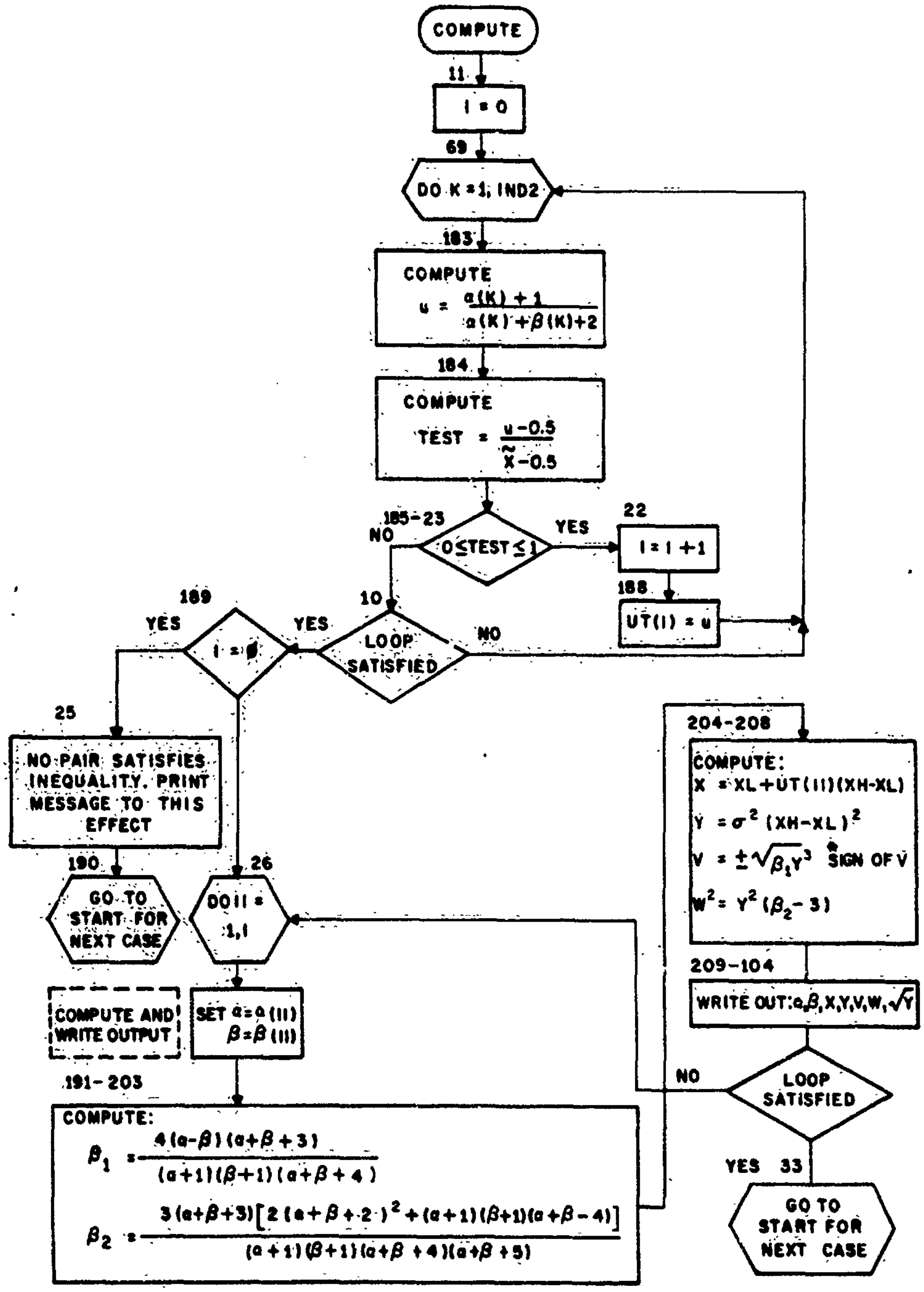



Figure 21. (concluded)

NOTE: The central range may not exceed

$$0.8\left(X_H - X_L\right) - 0.6 \left| 0.5\left(X_H - X_L\right) - \left(X_P - X_L\right) \right| .$$

If it does, the program will skip the case and go on to the next case. The beta distribution is computed as follows. First, the standard deviation $\sigma$ is determined. Compute

$$\tilde{X} = \frac{X_P - X_L}{X_H - X_L} .$$

$$C_{R_N} = \frac{C_R}{X_H - X_L} .$$

Let

$$U = \left| 0.5 - \tilde{X} \right| .$$

Then

$$\sigma = \frac{0.39\, C_{R_N}}{1 + A C_{R_N}^{\,a}}$$

where

$A = 0.148 - 0.603U + 3.98U^2$ , and

$a = 1.78 - 3.9U + 29.16U^2$ .

These equations were determined by curve-fitting, using several different values for the variables involved. Then determine

$$K = \frac{X_P - X_L}{X_H - X_P} .$$

The parameters $\alpha$ and $\beta$ are then determined by finding the real solution of the following* pair of equations (see Appendix II):

$$K = \alpha/\beta \quad ;$$

$$\sigma^2 = \frac{(\alpha+1)\,(\beta+1)}{(\alpha+\beta+2)^2\,(\alpha+\beta+3)}$$

For each pair of roots, compute

$$\mu = \frac{\alpha+1}{\alpha+\beta+2} \,.$$

Let

$$\tilde{X} = \frac{X_P - X_L}{X_H - X_L} = \frac{\alpha}{\alpha+\beta} \,,$$

and find all $\alpha$, $\beta$ which satisfy

$$0 \leq \frac{\mu - 0.5}{\tilde{X} - 0.5} \leq 1.$$

(In general, there will be only one such pair.)

When this pair $(\alpha, \beta)$ is determined, compute:

$$\beta_1 = \frac{4\,(\alpha+\beta)^2\,(\alpha+\beta+3)}{(\alpha+1)\,(\beta+1)\,(\alpha+\beta+4)^2} \,,$$

*The equation is solved in the equivalent form:

$$\sigma^2\,(K+1)^3\,\beta^3 + \left[7\sigma^2\,(K+1)^2 - K\right]\beta^2 + \left(16\sigma^2 - 1\right)(K+1)\,\beta + 12\sigma^2 - 1 = 0 \,.$$

and

$$\beta_2 = \frac{3(\alpha+\beta+3)\left[2(\alpha+\beta+2)^2+(\alpha+1)(\beta+1)(\alpha+\beta-4)\right]}{(\alpha+1)(\beta+1)(\alpha+\beta+4)(\alpha+\beta+5)}.$$

All necessary parameters are thus computed for determining the four moments.* These are computed as follows:

(a) first additive moment = mean = $X = X_L + \mu\left(X_H - X_L\right)$;

(b) second additive moment = (standard deviation)$^2$ = $Y = \left[\sigma\left(X_H - X_L\right)\right]^2$ ;

(c) third additive moment = $V = \pm\sqrt{\beta_1 Y^3} \ldots \begin{cases}\mu > X \Longrightarrow + \text{ sign} \\ \mu < X \Longrightarrow - \text{ sign}\end{cases}$ ; and

(d) fourth additive moment = $W = \left(\beta_2 - 3\right)Y^2$.

Also included in the output are $\alpha$, $\beta$ and $\sqrt{Y}$ (the standard deviation).

Program II

The analyst now has cost distributions (summarized by four moments) for each of the elements of the system under consideration. He specifies the additive-multiplicative combinations of these distributions that will yield the total system cost, and inserts the moments of each distribution into Program II (see Figure 22). Program II computes the specified combinations and prints out the moments of the total cost distribution, as well as certain other subtotals of interest to the analyst.

*In the main text, the four additive moments are referred to as $A_1$, $A_2$, $A_3$, and $A_4$; in this appendix they are referred to as X, Y, V, and W, respectively.

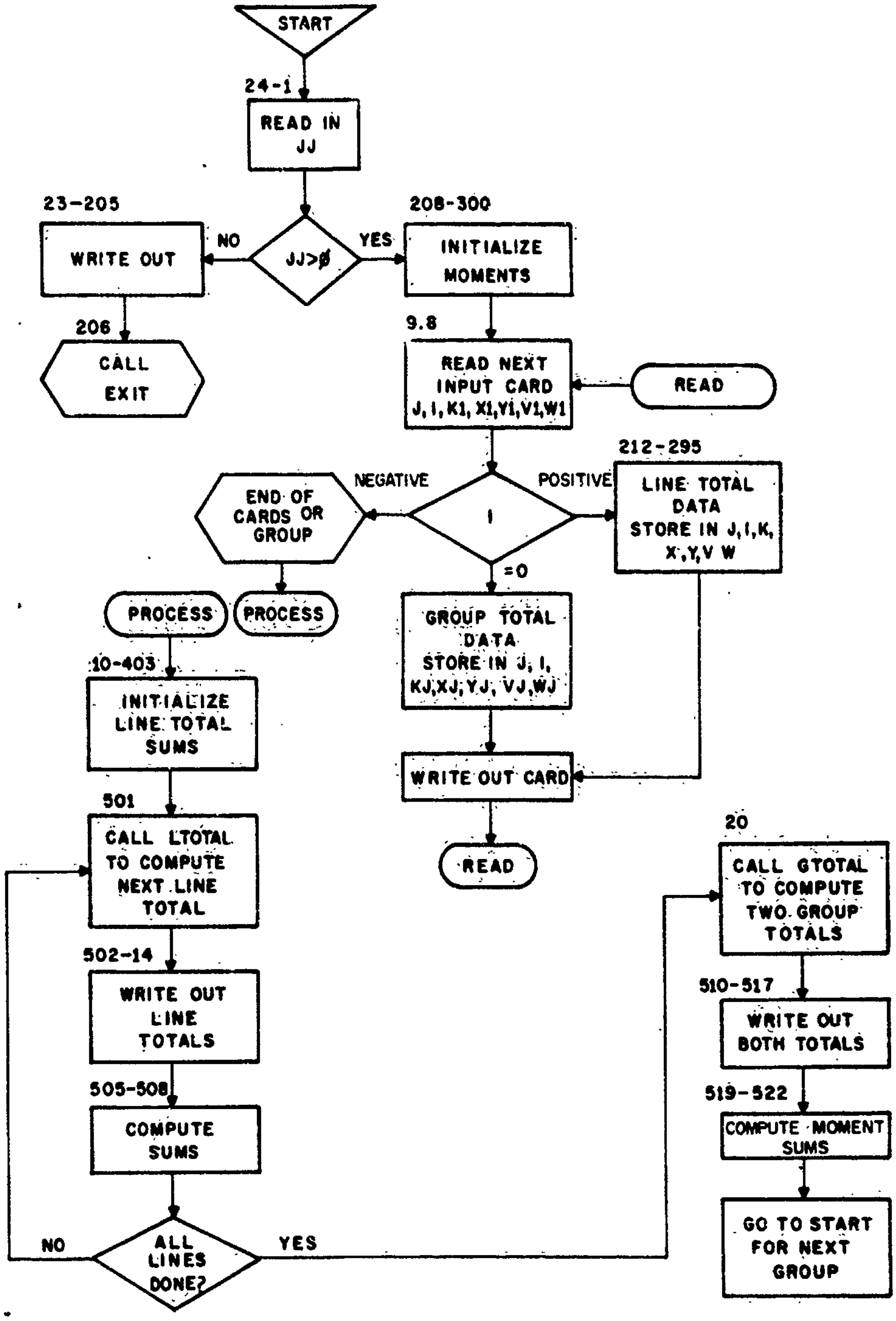



Figure 22. Program 11: Main Routine

Additive-Multiplicative Combinations of Independent Random Variables

Independent random variables $D_1$ and $D_2$ are given; the distributions for each summarized by its first four moments. Then

$$D_1 = D_1\left(X_1, Y_1, V_1, W_1\right) \text{ , and}$$

$$D_2 = D_2\left(X_2, Y_2, V_2, W_2\right) .$$

The distribution for $D_1 + D_2$ is computed by component adding moments on the basis of components; that is,

$$D_1 + D_2 = D_S\left(X_1 + X_2, Y_1 + Y_2, V_1 + V_2, W_1 + W_2\right) .$$

The distribution for $D_1D_2$ is computed by the following equations:*

$$D_1D_2 = D_M(X, Y, V, W),$$

where

$$X = X_1X_2 ;$$

$$Y = Y_1Y_2 + X_1^{\,2}\, Y_2 + X_2^{\,2}\, Y_1 ;$$

$$V = V_1V_2 + 3\left(V_1X_2Y_2 + V_2X_1Y_1\right) + V_1X_2^{\,3} + V_2X_1^{\,3} + 6X_1Y_1X_2Y_2; \text{ and}$$

$$W = W_1W_2 + 3\left(W_1Y_2^{\,2} + W_2Y_1^{\,2}\right) + 4\left(W_1X_2V_2 + W_2X_1Y_1\right)$$
$$+ 6\left(W_1X_2^{\,2}Y_2 + W_2X_1^{\,2}Y_1 + Y_1^{\,2}Y_2^{\,2}\right) + W_1X_2^{\,4} + W_2X_1^{\,4}$$

*This is the result of solving simultaneously the equations shown in Appendix III.

$$+ 12 \left( Y_1^{\,2} X_2 V_2 + Y_2^{\,2} X_1 V_1 + Y_1^{\,2} X_2^{\,2} Y_2 + Y_2^{\,2} X_1^{\,2} Y_1 + X_1 V_1 X_2 V_2 \right.$$

$$\left. + X_1 V_1 X_2^{\,2} Y_2 + X_2 V_2 X_1^{\,2} Y_1 \right) .$$

These formulas are equivalent to converting the additive moments into origin moments, multiplying these origin moments in terms of components and reconverting the product moments to additive moments.

Additive-multiplicative combinations of distributions may be computed by evaluating a composite term as in an algebraic expression. For example, to evaluate $D_1 \left(D_2 + D_3 D_4\right)$ , first evaluate $D_3 D_4$ , add $D_2$ to the result, and multiply the last result by $D_1$ . NOTE: This expression may <u>not</u> be evaluated in the form

$$D_1 D_2 + D_1 D_3 D_4 ,$$

since the two product terms are not independent (both depend on $D_1$) .

Additive-multiplicative combinations of distributions are specified for the program on two levels: lines and groups (see Figure 23).

<u>Lines</u>

A line is defined as an additive-multiplicative combination of distribution of the following form:

```
AAA {AA [AA (0+A0) + A A 0 + A 0 + 0 ] + A 0 + A 0 + A 0 + 0} ,
123  45  67  8 9 10  11 12 13  14 15  16    17 18  19 20  21 22  23
```

where each "A" or "0" symbol represents a distribution specified by its first four additive moments X, Y, V, W. The distribution for "A" is

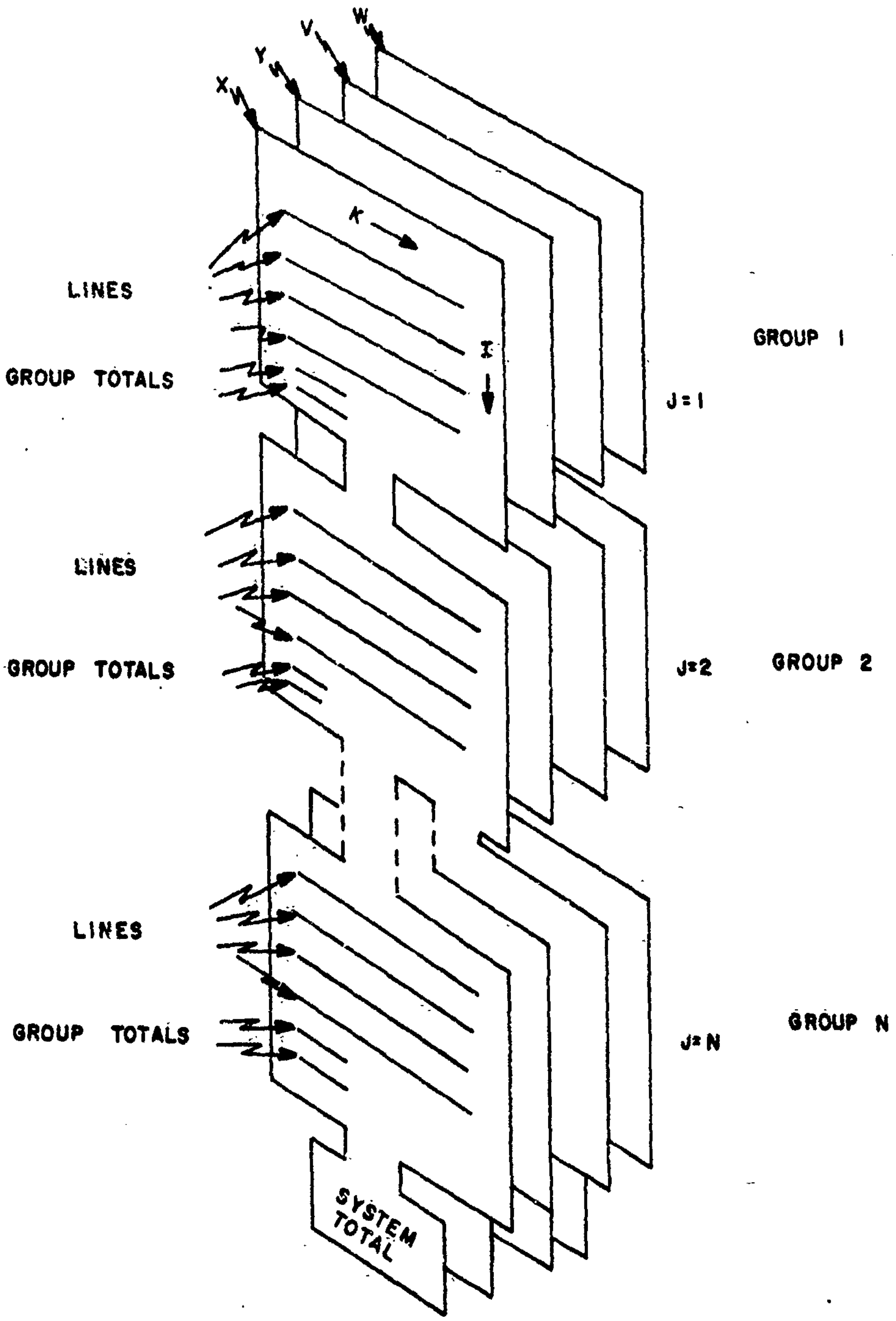



Figure 23. Generalized Cost Structure Built into Program 11

defined by the moments (1, ∅, ∅, ∅),* for "0" by the moments (∅, ∅, ∅, ∅). It may be observed that:

(a) if "0" is added to any distribution $D_1$ , the resulting moments are simply those of the distribution $D_1$ ; that is, $0 + D_1 = D_1$ ;

(b) if "0" is multiplied by any distribution $D_1$ , the resulting moments are (∅, ∅, ∅, ∅); that is, $0D_1 = 0$; and

(c) if "A" is multiplied by any distribution $D_1$ , the resulting distribution moments are simply those of $D_1$; that is, $AD_1 = D_1$ .

In other words, "0" and "A" resemble the ∅ and 1 of ordinary arithmetic. Because of this, the line total operator as it stands is equivalent to 0. If any "0" were replaced by some $D_1$, the line total would become $D_1$.

Lines are specified to be equivalent to certain additive-multiplicative combinations of distributions by the following technique: the analyst is permitted to replace any term in the line total operator by whatever distribution he desires. Such substitution will alter the value of the operator. For example, by replacing the 7th, 8th and 10th terms with the distribution $D_1$ , $D_2$ , and $D_3$ , respectively, the line will be made equivalent to the combination $D_1\left(D_2 + D_3\right)$. This can readily be seen by representing the operator symbolically:

$$1\cdot1\cdot1\left\{1\cdot1\left[1\cdot D_1\left(D_2+1\cdot D_3\right)+1\cdot1\cdot\emptyset+1\cdot\emptyset+\emptyset\right]+1\cdot\emptyset+1\cdot\emptyset+1\cdot\emptyset+\emptyset\right\}.$$

Application of the obvious rules reduces to this

$$1\left[D_1\left(D_2 + D_3\right)+\emptyset\right]+\emptyset = D_1\left(D_2+D_3\right) .$$

*The slashed zero (∅) is used to denote the numerical zero to avoid confusion.

Therefore, simply by specifying the line and the term into which a given set of distribution moments are to be substituted, a large variety of combinations can be evaluated.

Groups

On the second allowable level of combination of distributions, the various line totals are separated into groups, corresponding to subsystems of the system under consideration. First, all line moments for a group are computed. Then these moments are summed by components to produce a term S. This term may be combined with other distributions in the following operators:

$$\text{(G1)} \quad A_1 S(A_2 + A_3 0_4) \text{ , and}$$

$$\text{(G2)} \quad A_1 S(A_3 0_4) \text{ .}$$

Combinations are specified in the same manner used for lines: any A or 0 may be replaced by the moments of any distribution. A term replaced in the (G1) operator is also replaced in the (G2) operator by the same moments except for the second term, which appears only in the (G1) operator. It may be noted that if no term is replaced in the operators, the result of (G1) is S and the result of (G2) is "0."

Final Total

All moments obtained from the (G1) operator are added by components to give the final total. Thus, the (G1) group total should be so defined that the total system cost distribution is the sum of all of these totals.

## Program III

From Program II, the analyst has moment sets summarizing distributions for subsystems and total system cost uncertainty. These moment sets are now inserted into Program III (see Figure 24), which converts the moments into the beta distribution parameters and the bounds for the curve and for any specified percentage tails.

From the input moments, X, Y, V and W for each case, determine

$$\beta_1 = \frac{V^2}{Y^3} ; \qquad \beta_2 = \frac{W}{Y^2} + 3 .$$

The following equations are now to be solved for $\alpha$ and $\beta$, the beta curve parameters:

$$\beta_1 = \frac{4(\alpha-\beta)^2 (\alpha+ \text{[illegible]}}{(\alpha+1)(\beta+1)(\alpha+\beta+4)^2} ;$$

$$\beta_2 = \frac{3(\alpha+\beta+3)\left[2(\alpha+\beta+2)^2 + (\alpha+1)(\beta+1)(\alpha+\beta-4)\right]}{(\alpha+1)(\beta+1)(\alpha+\beta+4)(\alpha+\beta+5)} .$$

The constants $\beta_1$ and $\beta_2$ determine whether or not there is a meaningful solution to these equations.* (Refer to Appendix II for the details of the solution to these equations.) If there is a solution to these equations, use $\alpha$, $\beta$ to compute for the normalized beta curve:

$$\mu_N = \frac{\alpha+1}{\alpha+\beta+2} \qquad \text{(normalized mean)} ;$$

*For a unimodal beta distribution, $\alpha$ and $\beta$ must be positive and real.

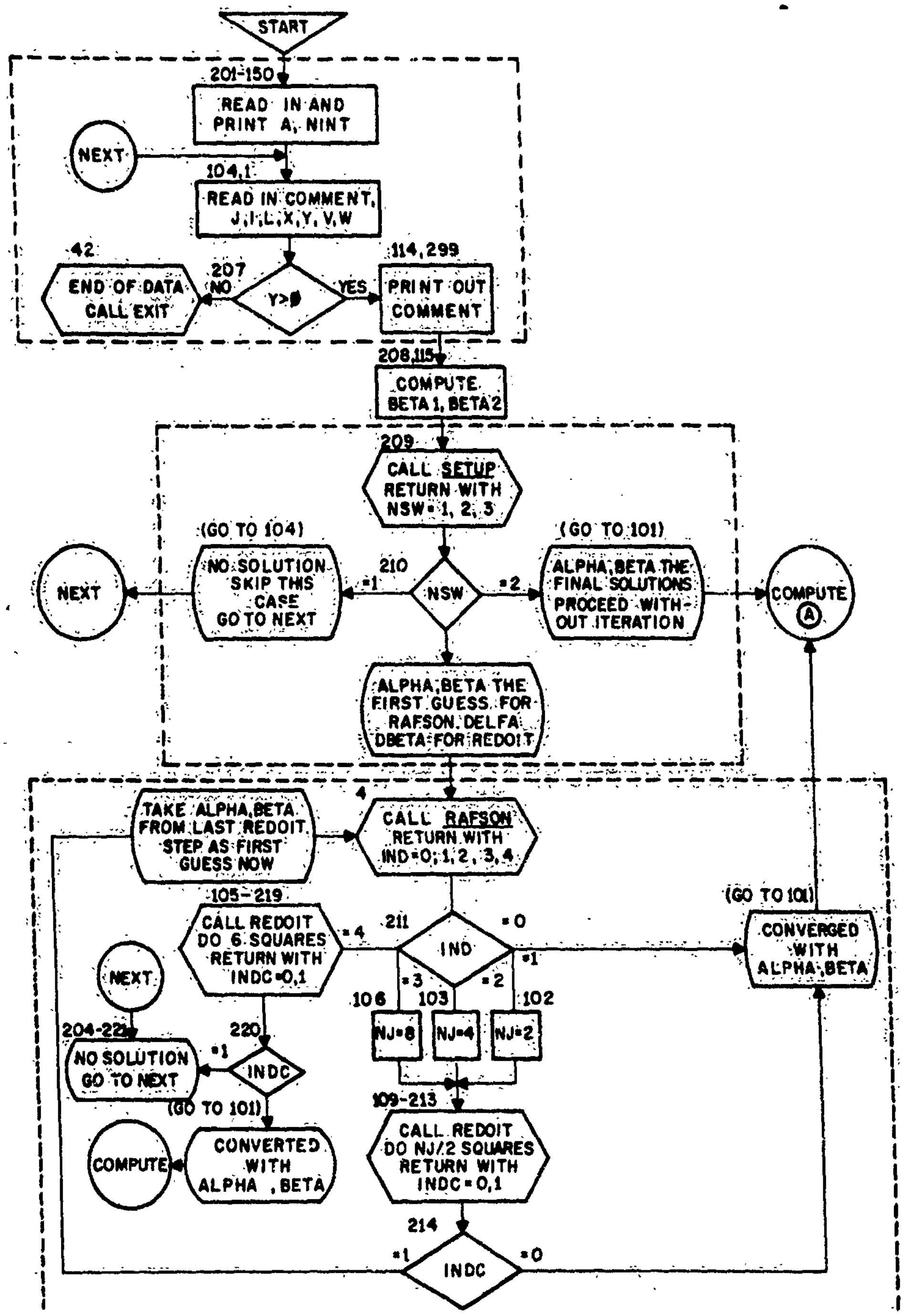



Figure 24. Program III: Main Routine

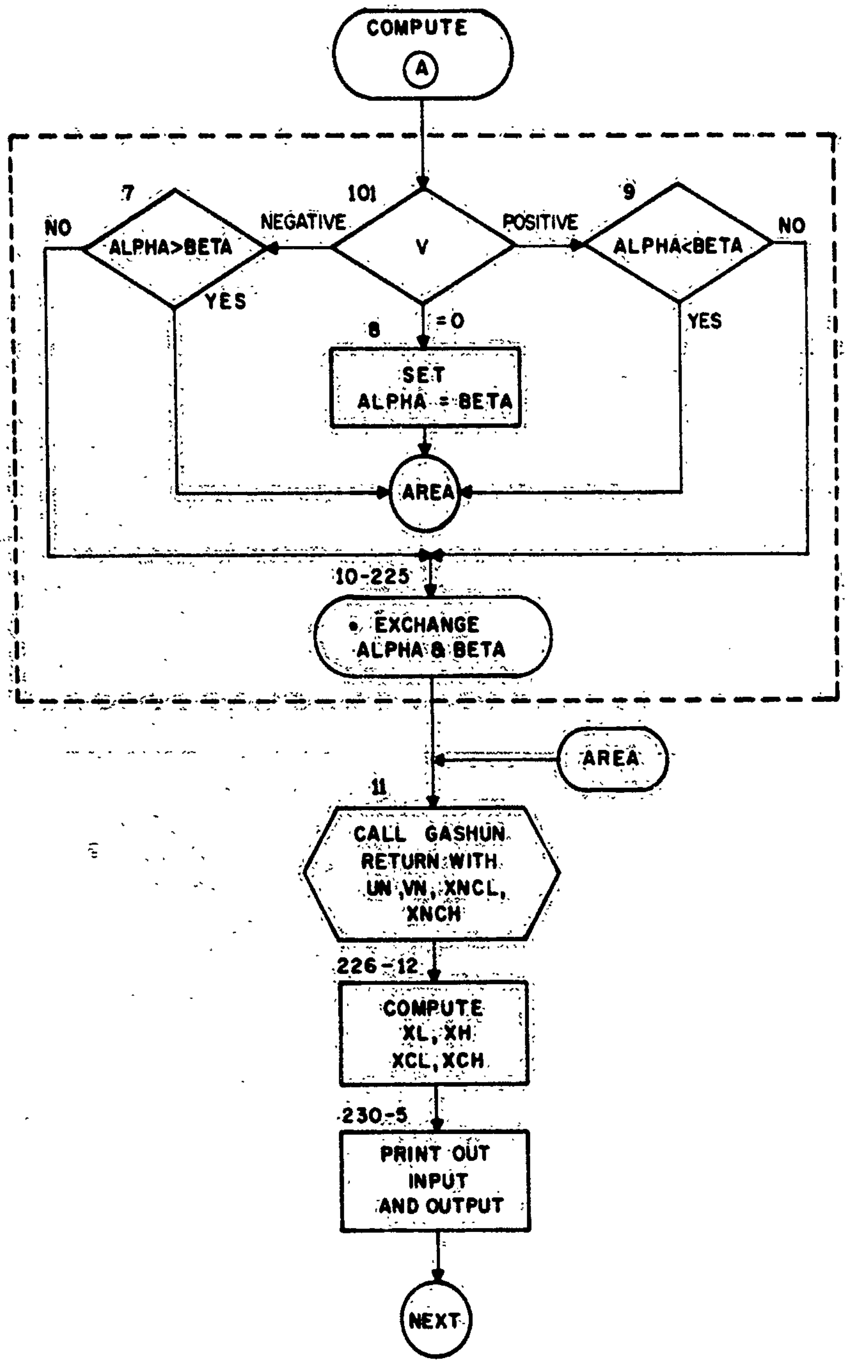



Figure 24. (concluded)

$$\sigma_N = \frac{\sqrt{(\alpha+1)\ (\beta+1)}}{(\alpha+\beta+2)\sqrt{\alpha+\beta+3}}$$ (normalized standard deviation); and

$$\tilde{X} = \frac{\alpha}{\alpha+\beta}$$ (normalized mode) .

The normalized curve must be integrated to determine the total area under the curve: Integrate

$$\text{AREA} = \int_{b_L}^{b_H} \left(\frac{X}{\tilde{X}}\right)^{\alpha} \left(\frac{1-X}{1-\tilde{X}}\right)^{\beta} dX$$

where

$$b_L = \max\left(0,\ \mu_N - 4\,\sigma_N\right)\ ;$$

$$b_H = \min\left(1,\ \mu_N + 4\,\sigma_N\right)\ .$$

This integration is performed numerically with Simpson's rule. The number of intervals used is NINT,* an input (800 is suggested). Determine the desired area under each tail:

$$\Delta A = A \times \text{AREA},$$

where

A is an input specifying the fractional tail desired.

*Number of intervals to be used in the Simpson integration.

The curve is integrated again, one interval at a time, from each end, until the X-values which are the bounds for the tails are reached. These are called XNCL and XNCH, respectively. From the bounds for the normalized curve, the bounds for the scaled beta curve are then determined. Compute:

$$\mu = X \qquad \text{(mean)};$$

$$\sigma = \sqrt{Y} \qquad \text{(standard deviation)} .$$

Then

the lower bound is $\quad XL = \mu - \dfrac{\sigma\mu_N}{\sigma_N}$ ;

the upper bound is $\quad XH = \mu + \dfrac{\sigma}{\sigma_N}\left(1-\mu_N\right)$ ,

the lower tail bound is $\quad XCL = \mu + \dfrac{\sigma}{\sigma_N}\left(XNCL - \mu_N\right)$ ; and

the upper tail bound is $\quad XCH = \mu + \dfrac{\sigma}{\sigma_N}\left(XNCH - \mu_N\right)$ .

## USE OF THE PROGRAMS

### Program I

#### Input

The input for Program I consists of the four parameters specifying the uncertainty in the cost of each individual system element:

$X_L$ = the lowest cost (lower 1 percent tail);
$X_H$ = the highest cost (upper 1 percent tail);

$X_P$ = the most probable cost (mode); and

$C_R$ = the length of the 80 percent central range (difference between upper and lower 10 percent tails).

The data for each element's cost uncertainty is punched on a separate input card. Each card has the following format (refer to Figure 25):

Column 1 is blank.

Columns 2 to 24 contain a Hollerith identification which the user may specify to identify the data. This identification will be printed out before the data for the case specified by the card when the program output is presented.

Columns 25 to 32 are left blank.

Columns 33 to 72 contain the four parameters $X_L$, $X_H$, $X_P$, and $C_R$. The data must be in E-format, contained in a field of ten columns. The first column of each field should be blank (columns 33, 43, 53, 63). The second contains the sign (if positive, it may be omitted). The third through sixth columns contain the three most significant digits of the number in the form X·XX. The seventh through tenth columns contain the power of 10 by which the significant digits must be multiplied to give the actual number. The exponent must be in the form E±XX (+ sign may be left as blank). Thus, the entire number is shown by:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ± | (X) | . | (X) | (X) | E | ± | (X) | (X) |

IMPORTANT: The last significant digit of the exponent <u>must</u> be in the tenth column.

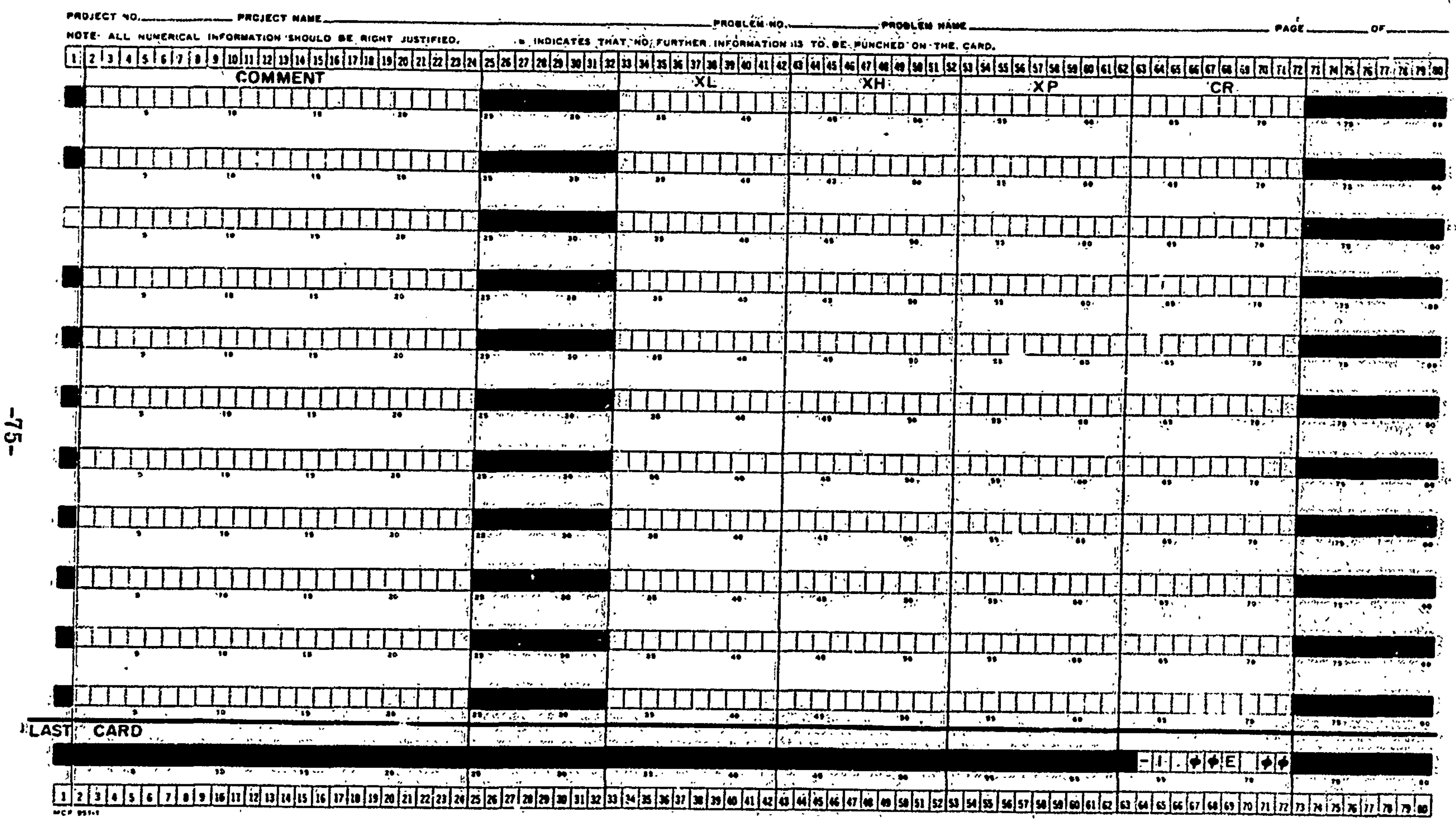



Figure 25. Input Format for Program 1

The cards may be arranged in any desired order; the cases will be processed in the order of their appearance in the deck. Following the card for the last (physical) case a special card with -1.ØØE ± ØØ punched in columns 44 to 53 must appear. This signals that all data has been processed, and the program will exit.

Program Deck

The program deck should have the following sequence of cards:

```
III   JOB
III   FMS
*     XEQ
      [ BINARY PROGRAM DECK ]
*     DATA
      [ DATA CARDS ]
      [ CR = -1.0 card ]
III   ENDJOB
```

Output

The output data (see Figure 26) will be arranged in the following manner. The identification will be printed at the head of the data for each case. The input data will then be printed; following this will come the output data:

X = the first additive moment (the mean) of the approximating beta distribution;

Y = the second additive moment (variance);

V = the third additive moment;

W = the fourth additive moment;

PROGRAM I

CASE-EXAMPLE I

| | X | Y | V | W | SQ.RT. Y | ALPHA | BETA |
|---|---|---|---|---|---|---|---|
| INPUT | XL 5.00E-01 | XH 3.00E 00 | XP 1.00E 00 | CR 1.20E 00 | | | |
| OUTPUT | 1.27E 00 | 2.06E-C1 | 5.26E-02 | -1.24E-02 | 4.54E-01 | 6.95E-01 | 2.78E 00 |

CASE-SSD 4th ENG/MAT 2.11

| | X | Y | V | W | SQ.RT. Y | ALPHA | BETA |
|---|---|---|---|---|---|---|---|
| INPUT | XL 2.00E 06 | XH 2.00E 07 | XP 5.00E 06 | CR 1.00E 07 | | | |
| OUTPUT | 7.91E 06 | 1.40E 13 | -2.42E-20 | -9.10E 25 | 3.74E 06 | 3.53E-01 | 1.77E 00 |

CASE-SSD 4th RE/800S 2.12

| | X | Y | V | W | SQ.RT. Y | ALPHA | BETA |
|---|---|---|---|---|---|---|---|
| INPUT | XL 2.00E 05 | XH 1.00E 06 | XP 3.00E 05 | CR 4.50E 05 | | | |
| OUTPUT | 4.55E 05 | 2.86E 10 | 2.82E 15 | -3.58E 20 | 1.69E 05 | 2.33E-01 | 1.63E 00 |

CASE-SSD 4th UNIT 2.14

| | X | Y | V | W | SQ.RT. Y | ALPHA | BETA |
|---|---|---|---|---|---|---|---|
| INPUT | XL 1.10E 05 | XH 2.50E 05 | XP 1.50E 05 | CR 7.00E 04 | | | |
| OUTPUT | 1.61E 05 | 6.87E 08 | 6.98E 12 | -2.36E 17 | 2.62E 04 | 1.03E 00 | 2.56E 00 |

CASE-SSD 4th L/SI 2.13N

| | X | Y | V | W | SQ.RT. Y | ALPHA | BETA |
|---|---|---|---|---|---|---|---|
| INPUT | XL 4.50E 02 | XH 9.00E 02 | XP 6.00E 02 | CR 3.00E 02 | | | |
| OUTPUT | 6.47E 02 | 1.19E 04 | 2.56E 05 | -1.31E 08 | 1.09E 02 | 3.95E-01 | 7.90E-01 |

Figure 26. Sample Output for Program 1

$\left.\begin{array}{l}\bar{\alpha}\\ \beta\end{array}\right\}$ = the parameters which determine the shape of the beta curve; and

$\sqrt{Y}$ = the standard deviation

All data will be printed in E-format to three significant decimal digits.

Program II

It is important that the user read the previous discussion of Program II for an understanding of how to use this program.

Input

Each group is processed separately by the program, so all cards relating to the same group must be placed together. For each group, the following input cards must be present (see Figure 27):

The first card in each set of group cards must have a positive number JJ (preferably the number of the group) in Column 5. (If it is a two-digit number, use columns 4 and 5.)

The next set of cards specifies the line data for this group. There is one card for each set of moments which are to be used in this group. Each card has an identification field in columns 2 to 24 as in Program I.

Column 25 is blank.

Columns 26 and 27 contain the group number, J, for the present group. All cards for one group must be placed together.

Columns 28 and 29 contain the line number, I, in which this particular set of moments is to be combined. The I's in each group must be

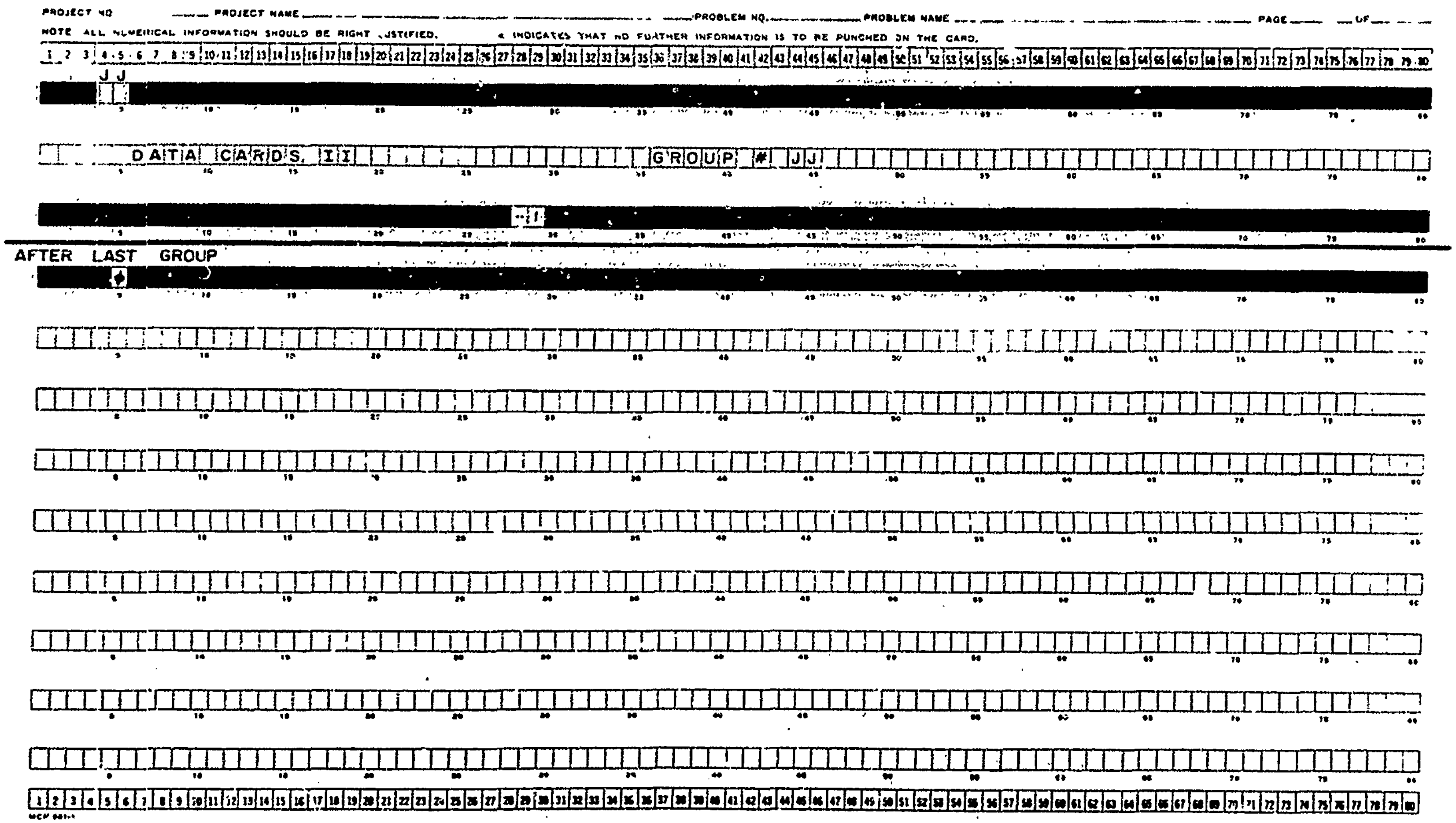



Figure 27.. Input Format for Program 11

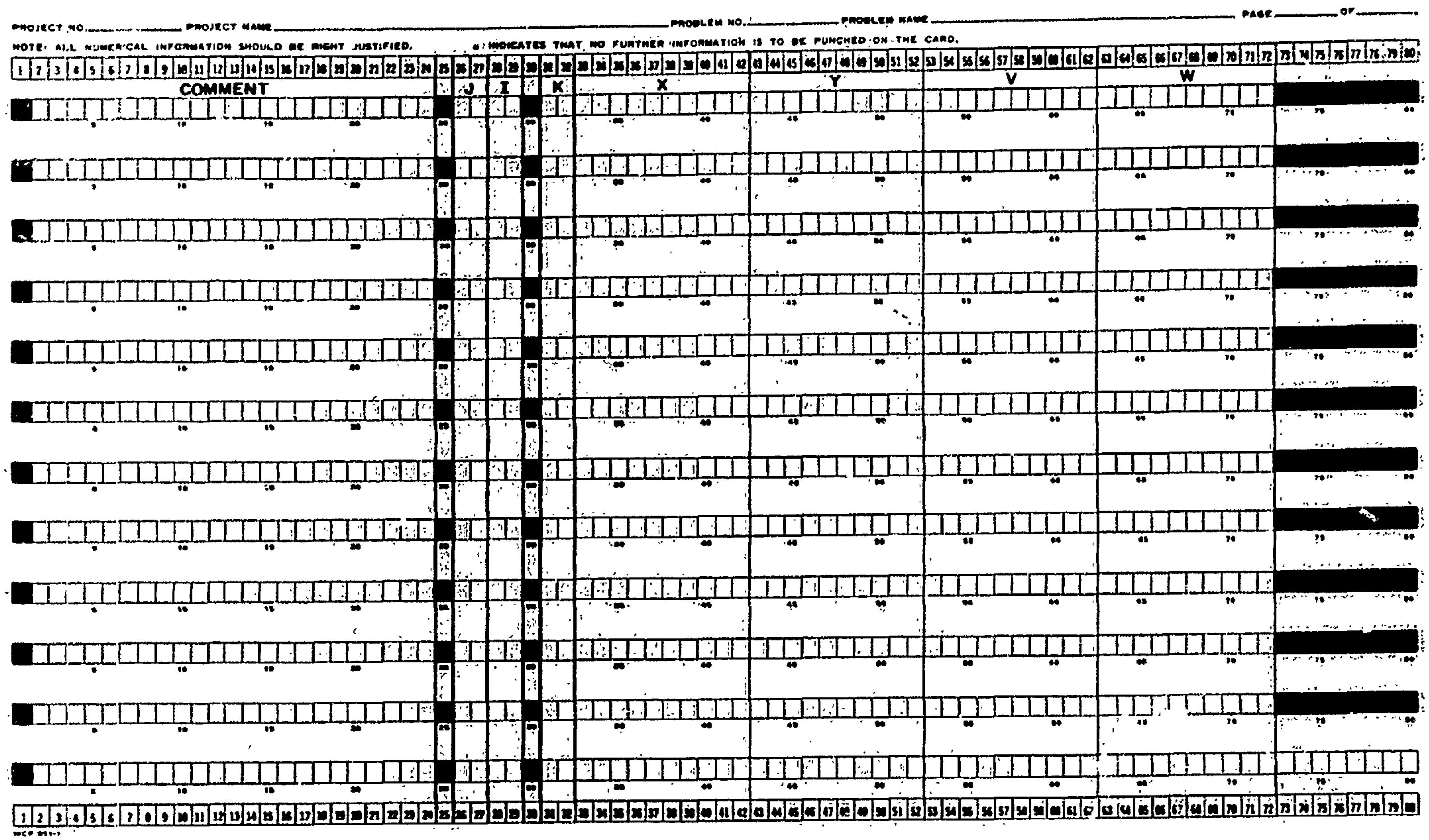



Figure 27. (concluded)

arranged from 1 consecutively up to IMAX, which cannot be greater than 5∅.

Column 30 is blank.

Columns 31 and 32 contain the term (column) number K for which the moments on this card are to be substituted in the line total specified by I. The K term may be any integer between 1 and 23, inclusive.

Columns 33 to 72 contain the moments X, Y, V, W which are to be substituted for term K in Line I in group J. These columns have the same format as on the Program I data cards.

The group total cards follow the line cards. These have the same format as the line cards, except that:

I is always ∅ for a group total input; and
K is the term number for which the moments on the card will be substituted (K ranges from 1 to 4, inclusive).

A card which is blank, except for a "-1" in columns 28 and 29, follows the last group total card (or last line total card if there is no group total data). This signals the end of the group.

The cards for all subsequent groups are arranged in the same way. A card which is blank except for a ∅ punched in column 5 appears after the I = -1 card for the last group. This signals the end of the data.

<u>Program Deck</u>

III JOB

III FMS

* ZEQ

[BINARY DECK]

* DATA

[JJ = 1] card

[Line Total Data I > Ø]

[Group Total Data I = Ø]

[I = -1] card

[JJ = 2] card

etc.

[I = -1] card

[JJ = Ø]

III ENDJOB

<u>Output</u>

The line total input data for the first physical group in the deck appears on the first page followed by the group total input, if any, for this group on the next page (see Figure 28). Each entire card will be printed out except for the JJ and I = -1 cards.

The line total moments X, Y, V, W for this group are shown on the next page. In addition, three characters will be printed:

J = the group number;

I = the line number; and

L = Ø, to indicate a line total.

The X, Y, V, W are interpreted in the same way as for Program I, except that now they are moments representing the distributions of random variables which are aggregations of other system element costs.

The group total output for the group appears on the next page.

PROGRAM II

INPUT TO CALCULATE LINE MOMENTS

| | J | I | K | X | Y | V | W |
|---|---|---|---|---|---|---|---|
| SD 4th ENG/MAT 2.11 | 1 | 1 | 8 | 7.91E 06 | 1.40E 13 | -2.42E-20 | -9.10E 25 |
| MULTIPLIER | 1 | 2 | 8 | 1.00E 00 | 0. | 0. | 0. |
| SSD 4th RE/BOOS 2.12 | 1 | 2 | 22 | 4.55E 05 | 2.86E 10 | 2.82E 15 | -3.58E 20 |
| SSD 4th UNIT 2.14 | 1 | 2 | 23 | 1.61E 05 | 6.87E 08 | 6.98E 12 | -2.36E 17 |
| MULTIPLIER | 1 | 2 | 1 | 5.00E 00 | 0. | 0. | 0. |
| SSD 4th L/S1 2.13N | 1 | 2 | 5 | 6.47E 02 | 1.19E 04 | 2.56E 05 | -1.31E 06 |
| SSD 4th L/S2 2.13M | 1 | 2 | 6 | 9.60E-01 | 1.29E-02 | -5.97E-04 | -5.76E 05 |
| SSD 4th L/S3 2.13C | 1 | 2 | 7 | 2.00E 03 | 1.12E 04 | 0. | -4.31E 08 |
| SSD 4th L/S4 2.13P | 1 | 2 | 10 | 3.61E-01 | 5.52E-03 | 4.72E-05 | -2.38E 05 |

OUTPUT FOR LINE MOMENTS

| J | I | L | X | Y | V | W |
|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 7.91E 06 | 1.40E 13 | -2.42E-20 | -9.10E 25 |
| 1 | 2 | 0 | 1.15E 07 | 4.22E 12 | 2.92E 18 | -2.62E 24 |

OUTPUT OF GROUP MOMENTS

| J | I | L | X | Y | V | W |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1.94E 07 | 1.82E 13 | 2.92E 18 | -9.36E 25 |
| 1 | 0 | 2 | 0. | 0. | 0. | -0. |

OUTPUT OF TOTAL COST MOMENTS

| J | I | L | X | Y | V | W |
|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 1.94E 07 | 1.82E 13 | 2.92E 18 | -9.36E 25 |

Figure 28. Sample Output for Program 11

J = group number

I = ∅, to indicate group total;

L = 1, for the (G1) total; and

L = 2, for the (G2) total

All groups will have their output in this form. After the last group total output the final total will appear. Its characters are

$$J = 0; I = 0; L = 1.$$

All output data will be punched on cards for direct insertion into Program III.

Program III

Input

The first card specifies the two parameters A and NINT for the integration process (see Figure 29):

Columns 1 to 5 are blank.

Columns 6 to 14 contain A, the fraction of the total area desired in each tail.

Columns 15 to 35 are blank.

Columns 36 to 38 contain NINT.

Rest of card blank.

The next cards contain the data to be processed by the program.

Columns 2 to 24 contain identification.

Column 25 is blank.

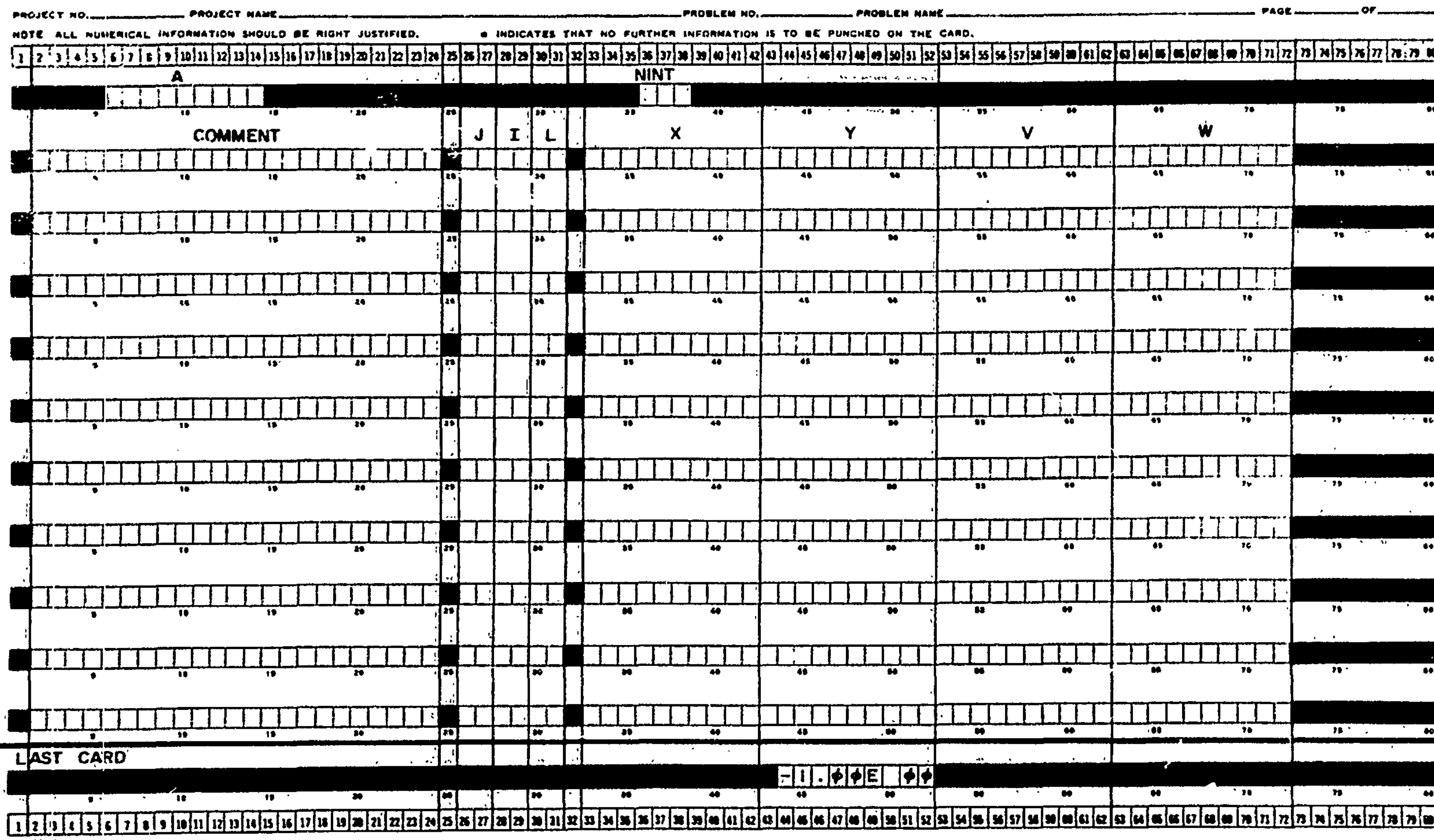



Figure 29. Input Format for Program 111

Columns 26 and 27, 28 and 29, 30 and 31, and J, I, and L indicate the control characters printed with Program II output. These are merely for the purpose of identification; they are not used by the program.

Column 32 is blank.

Columns 33 to 72 X, Y, V, W are in same format as Programs I and II.

A card with nine columns (44 to 52) containing a -1.ØØE+ØØ must appear after these cards. This signals the end of the data.

Program Deck

```
III  JOB
III  FMS
*    XEQ
     [PROGRAM DECK]
*    DATA
     [DATA CARDS]
     [Y = 1.Ø   card]
III  ENDJOB
```

Error Returns

A number of conditions will cause the program to indicate that no solution can be found for the present case. When this happens, the program will proceed with the next data, because the distribution specified by the moments for this case cannot be converted by the program.

Output

The output data (see Figure 30) for each case will be preceded by the identification field and the input data. The output will consist of:

$\left.\begin{matrix}\alpha\\ \beta\end{matrix}\right\}$, the parameters* determining the shape of the beta curve;

$\left.\begin{matrix}X_L\\ X_H\end{matrix}\right\}$, the bounds* for the linearly scaled curve; and

$\left.\begin{matrix}XC_L\\ XC_H\end{matrix}\right\}$, the bounds for the 100A percent tails.

*Refer to the definition of the linearly scaled beta distribution in the Glossary; $a = X_L$ and $a+b = X_H$.

10.00 PERCENT TAILS    800 INTERVALS USED IN NUMERICAL INTEGRATION.    PROGRAM III

CASE----- EXAMPLE 1
INPUT, X = 2.54E 01    Y = 412E 00    V = 1.05E 00    W = -2.48E-01

| | JIL | ALPHA | BETA | XL | XH | XCL | XCH |
|---|---|---|---|---|---|---|---|
| EXAMPLE 1 | 202 | 4.78E 01 | 1.04E 02 | 8.18E 00 | 6.25E 01 | 2.28E 01 | 2.80E 01 |

CASE--- 4th STAGE VEHICLE SUB.
INPUT, X = 1.94E 07    Y = 1.82E 13    V = 2.92E 18    W = -9.36E 25

| | JIL | ALPHA | BETA | XL | XH | XCL | XCH |
|---|---|---|---|---|---|---|---|
| 4th STAGE VEHICLE SUB. | 102 | 4.42E 00 | 7.57E 00 | 6.27E 06 | 4.02E 07 | 1.40E 07 | 2.51E 07 |

Figure 30. Sample Output for Program 111

## APPENDIX VII

## DETAILS OF FORTRAN SUBROUTINES

### PROGRAM I

Program I requires the solution of a cubic equation. The routine CUBERT (see Figure 31) is used for this purpose.

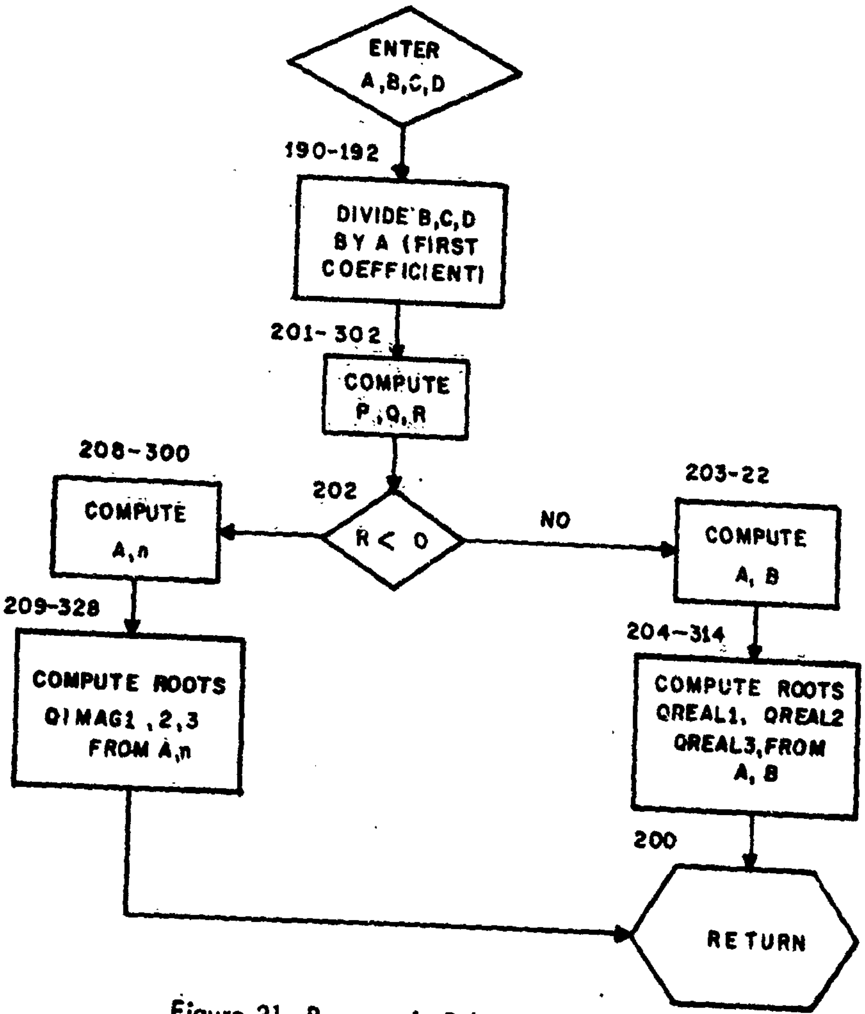



Figure 31. Program 1: Subroutine CUBERT

PROGRAM III

Subroutine SETUP

The subroutine SETUP (see Figure 32) determines whether a solution exists to the equations in Program III for $\beta_1 (\alpha, \beta)$ and $\beta_2 (\alpha, \beta)$. It checks certain criteria on $\beta_1$ and $\beta_2$ to determine the nature of the equations for $\alpha$ and $\beta$ as shown below.

1) The program will skip any case for which $\beta_1$ is greater than 4 or $\beta_2$ is greater than 9.

2) If $\beta_1$ is less than 0.ØØ1, no iteration is necessary. In this case, the following checks are made:

(a) If $\beta_2$ is less than 2.9994, then the final solution is

$$\alpha = \beta = \frac{5\beta_2 - 9}{2(3 - \beta_2)} \quad .$$

(b) If $\beta_2$ is between 2.9994 and 3.0015, the final solution is approximately

$$\alpha = \beta = 5ØØØ.$$

These are the answers used in this case.

(c) If $\beta_2$ is greater than 3.0015, there is no solution, and this case will be skipped.

3) If $\beta_1 > 0.ØØ1$, $\beta_2$ is compared to 3.ØØ1 with the following results:

(a) If $\beta_2$ is less than 3.ØØ1, iteration is necessary, and procedure (4) is used.

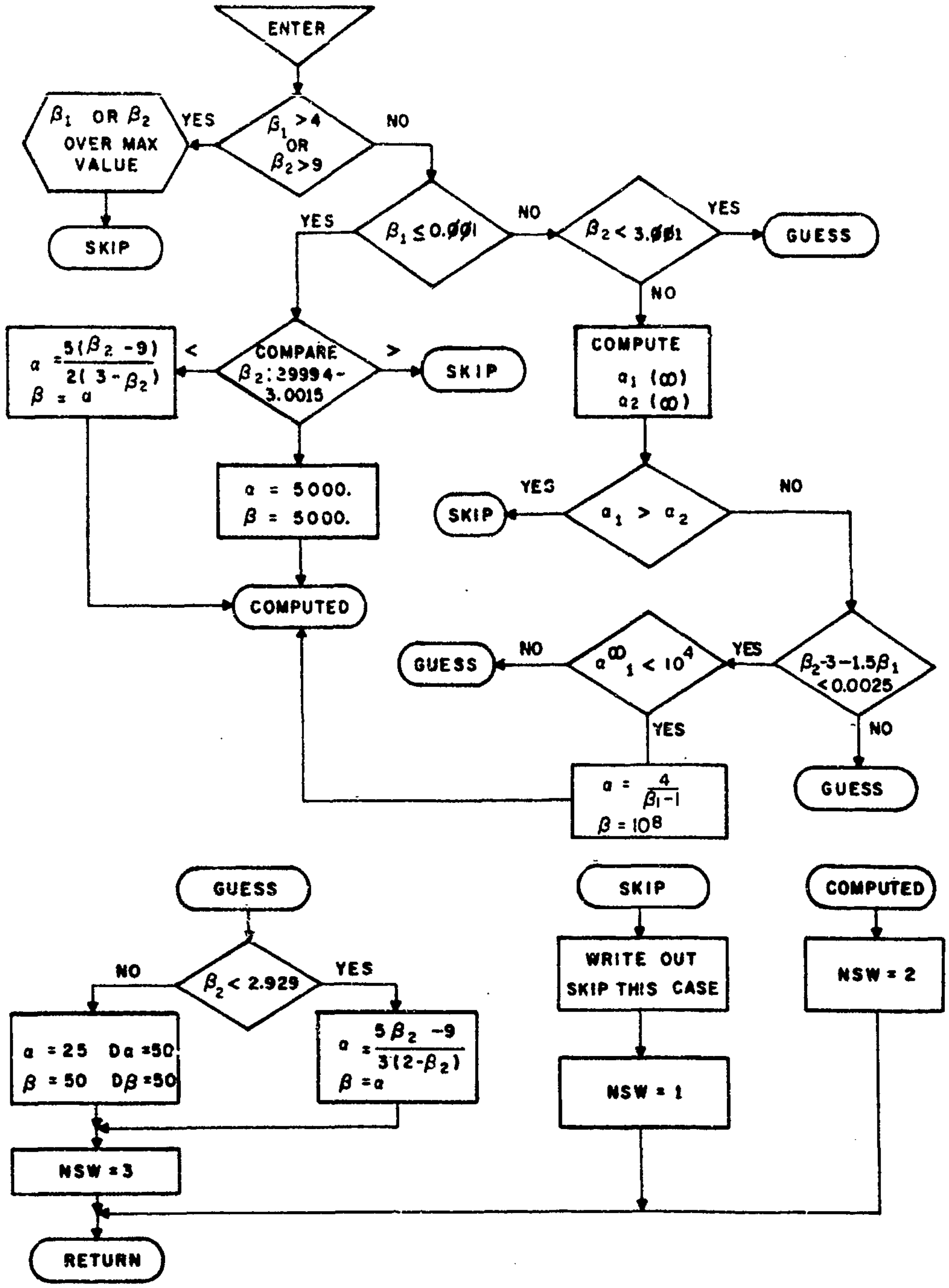



Figure 32. Program 111: Subroutine SETUP

(b) If $\beta_2$ is greater than 3.001, then the trial solution for $\beta$ is $+\infty$. In this case, equations are derived for $\beta_1$ and $\beta_2$ with $\beta = \infty$, and the value of $\alpha$ determined in each equation. The equations are:

$$\alpha_\infty\left(\beta_1\right) = \frac{4}{\beta_1} - 1$$

$$\alpha_\infty\left(\beta_2\right) = \frac{9 - \beta_2}{\beta_2 - 3}$$

If $\alpha_\infty\left(\beta_1\right) > \alpha_\infty\left(\beta_2\right)$, there is no solution possible, and the case will be skipped.

If $\alpha_\infty\left(\beta_1\right) \leq \alpha_\infty\left(\beta_2\right)$, then the number $\left| \beta_2\left[\alpha=\alpha_\infty \; \beta_1 \;, \beta=\infty\right] - \beta_2 \right|$ = beta is computed. Should $\beta \leq 0.0025$ and $\alpha_\infty\left(\beta_1\right) \leq 10^4$, then this $\alpha_\infty \; \beta_1$ is the solution for $\alpha$, and $10^8$ is used for $\beta$.

If either of these conditions is not satisfied, iteration is used, and the program proceeds according to the next step.

4) When iteration is used, the initial guess for $\alpha$ and $\beta$ is determined thus:

(a) If $\beta_2 \leq 2.929$, then the guess for $\beta$ is

$$\beta = \frac{5\beta_2 - 9}{2\left(3 - \beta_2\right)} \;;$$

$\alpha = \beta/2$ is used for the RAFSON routine and $\alpha = \beta$ for REDOIT.

(b) If $\beta_2 > 2.929$, then $\alpha = \beta = 50$ is used for REDOIT, and $\alpha = 25$, $\beta = 50$ for RAFSON routine.

If SETUP determines that an iteration is necessary, the solution of these equations is obtained through an iteration method using an alternating procedure between the Newton-Rafson method and binary search in two dimensions.

Twelve iteration steps are first attempted with the Newton-Rafson method; the initial guess on $\alpha$ and $\beta$ is either specified by the user or is determined internally if the user so desires. If this method fails to converge, a closer approximation is obtained for the initial guess using one step of the binary search, and twelve more steps are performed with the Rafson method. If this still fails, another closer initial guess is determined with two steps in the binary search, and twelve more Newton-Rafson steps are taken. Finally, if convergence is still not attained, ten steps of the binary search are taken.

Subroutine RAFSON

This subroutine (see Figure 33) iterates for the solution to the equations

$$\beta_1 = \frac{4(\alpha-\beta)^2\,(\alpha+\beta+3)}{(\alpha+1)(\beta+1)(\alpha+\beta+4)^2}\ ;$$

$$\beta_2 = \frac{3(\alpha+\beta+3)\left[2\,(\alpha+\beta+2)^2 + (\alpha+1)\,(\beta+1)\,(\alpha+\beta-4)\right]}{(\alpha+1)(\beta+1)\,(\alpha+\beta+4)\,(\alpha+\beta+5)}\ ;$$

by the Newton-Rafson method.

An initial guess is made by the SETUP routine, with this $(\alpha, \beta)$, the first partial derivatives of these equations are computed:

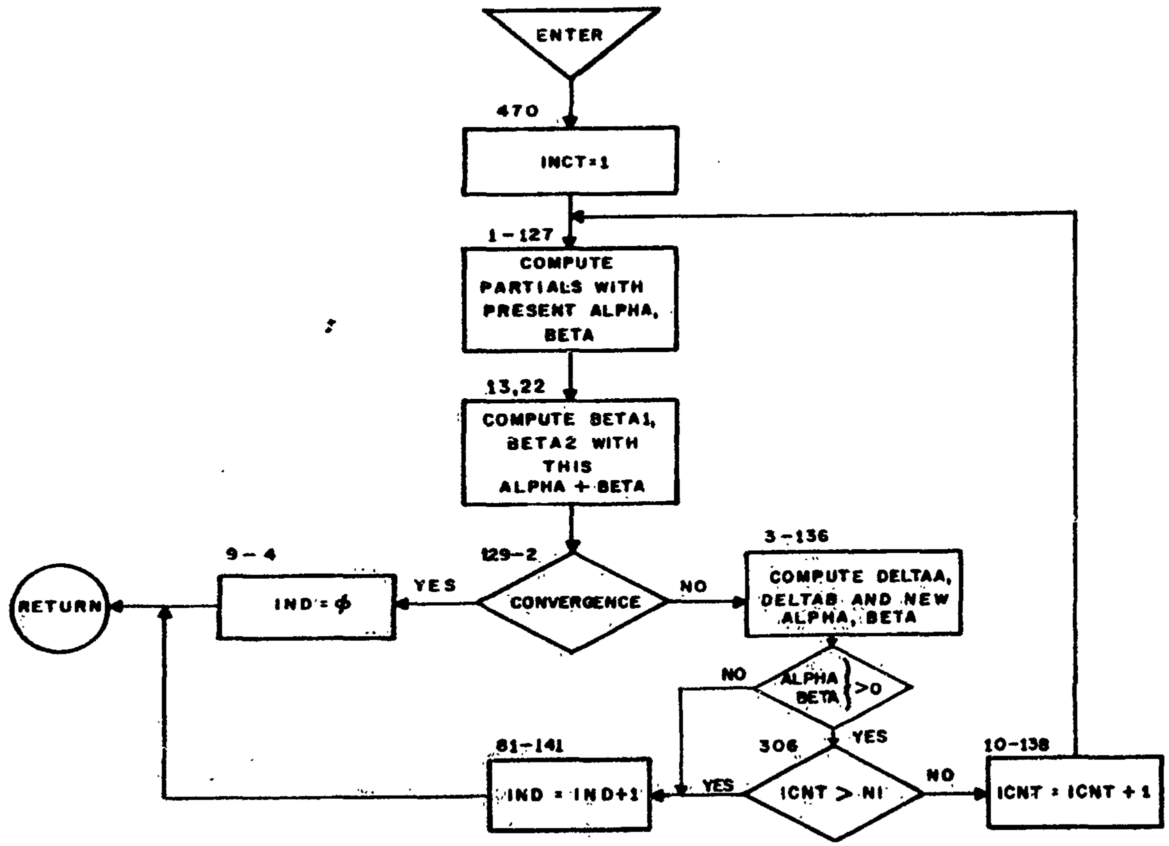



Figure 33. Program 111: Subroutine RAFSON

$$\frac{\partial \beta_1}{\partial \alpha} = \frac{4(\alpha-\beta)(3\alpha+\beta+6)(\beta^2+\alpha\beta+2\alpha+4\beta+4)}{(\alpha+1)^2(\beta+1)(\alpha+\beta+4)^3}$$

$\dfrac{\partial \beta_1}{\partial \beta} = \dfrac{\partial \beta_1}{\partial \alpha}$ , with $\alpha$ and $\beta$ interchanged;

$$\frac{\partial \beta_2}{\partial \alpha} = \frac{3(\alpha+1)(\alpha+\beta+4)(\alpha+\beta+5)\left[2(\alpha+\beta+2)^2+(\alpha+1)(\beta+1)(\alpha+\beta-4)+(\alpha+\beta+3)(\beta^2+2\beta+2\alpha\beta)+(6\alpha+5)\right]-3(\alpha+\beta+3)\left[2(\alpha+\beta+2)^2+(\alpha+1)(\beta+1)(\alpha+\beta-4)\right]}{(\alpha+1)^2(\beta+1)(\alpha+\beta+4)^2(\alpha+\beta+5)^2} \times$$

$$\frac{(3\alpha^2+4\alpha\beta+20\alpha+\beta^2+11\beta+29)}{(\alpha+1)^2(\beta+1)(\alpha+\beta+4)^2(\alpha+\beta+5)^2},$$

$\dfrac{\partial \beta_2}{\partial \beta} = \dfrac{\partial \beta_2}{\partial \alpha}$ , with $\alpha$ and $\beta$ interchanged.

With these, the increments for $\alpha$ and $\beta$ are determined using the relations of the Newton-Rafson method:

$$\Delta\alpha = -\frac{\begin{vmatrix} \beta_1(\alpha,\beta)-\beta_1 & \dfrac{\partial \beta_1}{\partial \beta} \\ B_2(\alpha,\beta)-\beta_2 & \dfrac{\partial \beta_2}{\partial \beta} \end{vmatrix}}{\begin{vmatrix} \dfrac{\partial \beta_1}{\partial \alpha} & \dfrac{\partial \beta_1}{\partial \beta} \\ \dfrac{\partial \beta_2}{\partial \alpha} & \dfrac{\partial \beta_2}{\partial \beta} \end{vmatrix}},$$

and

$$\Delta\beta = - \frac{\beta_2(\alpha, \beta) - \beta_2 - \partial\beta_2/\partial\alpha \;\; \Delta\alpha}{\partial\beta_2/\partial\beta} .$$

Then new values of $\partial$ and $\beta$ are computed by

$$\alpha = \alpha + \Delta\alpha , \; \beta = \beta + \Delta\beta \;\; ,$$

and $\beta_1(\alpha, \beta)$, $\beta_2(\alpha, \beta)$ are computed. If these do not satisfy the criterion of 0.01 percent, the process is repeated up to a maximum of twelve times each time it is called.

Subroutine REDOIT

In the event that Newton-Rafson iteration fails to converge, this subroutine (see Figure 34) is used to compute a guess for RAFSON that will be closer than the last. The routine is also used to attain convergence if overflow or divide checks makes RAFSON diverge.

The procedure is a binary search in two dimensions. An initial guess is computed by SETUP. Each time after the first, the last value is used as the guess.

The initial $\beta$ is held fixed, and $\alpha$ is incremented by 35 percent, until the last two values of $\beta_1(\alpha, \beta)$ fall on opposite sides of $\beta_1$ . (If after 20 times no interval containing the solution exists, the program exits and skips this case.)

If an interval is found successive, linear interpolations are performed until the computed $\beta_1$ is within 0.1 percent of the desired value (maximum of 10 times). This completed, the same thing is done with $\beta$ , holding the newly computed $\alpha$ fixed until $\beta_2$ is satisfied.

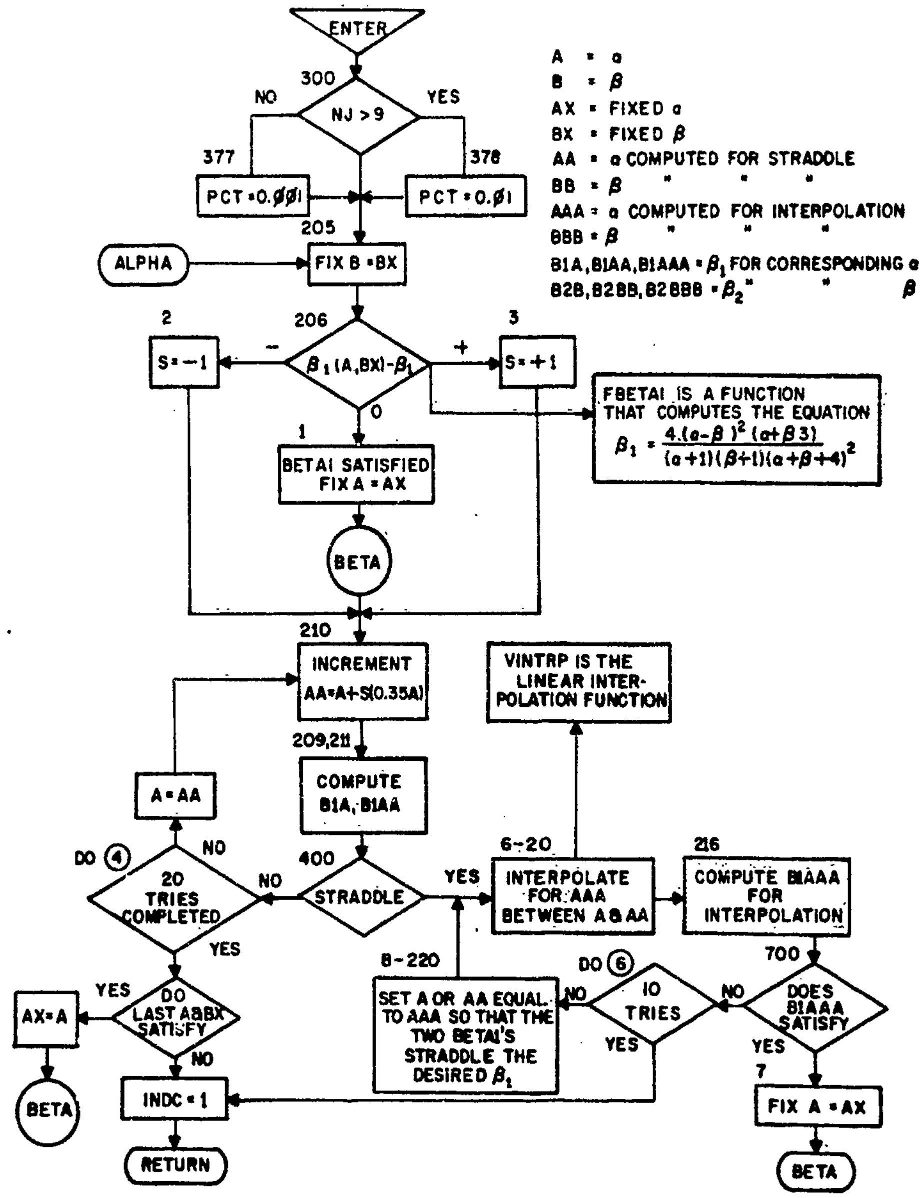



Figure 34. Program 111: Subroutine REDOIT

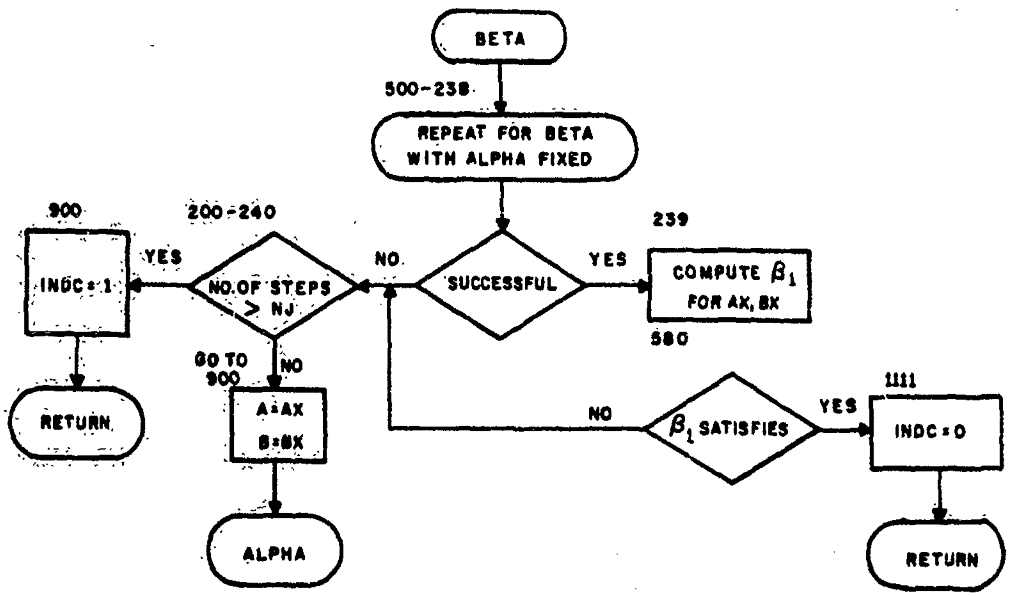



Figure 34. (concluded)

With the new $(\alpha, \beta)$, $\beta_1(\alpha, \beta)$ is computed. If the results are within 0.1 percent of $\beta_1$, set the indicator to Ø and return; if not, the process is repeated. The number of times the process is repeated is a function of the number of times RAFSON has been used:

| | | |
|---|---|---|
| 1 time | - | 2 times |
| 2 times | - | 4 times |
| 3 times | - | 8 times |
| 4 times | - | 12 times |

NOTE: On the last trial for REDOIT, a 1 percent criterion instead of 0.1 percent is used. If convergence is not attained after the maximum number of steps, set the indicator to 1 and return.

The routine GASHUN performs the computations for the normalized beta curve.

## Subroutine GASHUN

This subroutine (see Figure 35) computes the mean, standard deviation, and tail end points for the normalized beta curve. The mean

$$\mu_N = \frac{\alpha+1}{\alpha+\beta+2} ,$$

and standard deviation

$$\sigma_N = \frac{\sqrt{(\alpha+1)(\beta+1)}}{(\alpha+\beta+2)\sqrt{\alpha+\beta+3}}$$

are computed.

Then the following integral

$$\int_{b_L}^{b_H} \left(\frac{X}{\tilde{X}}\right)^{\alpha} \left(\frac{1-X}{1-\tilde{X}}\right)^{\beta} dX ,$$

is to be evaluated;

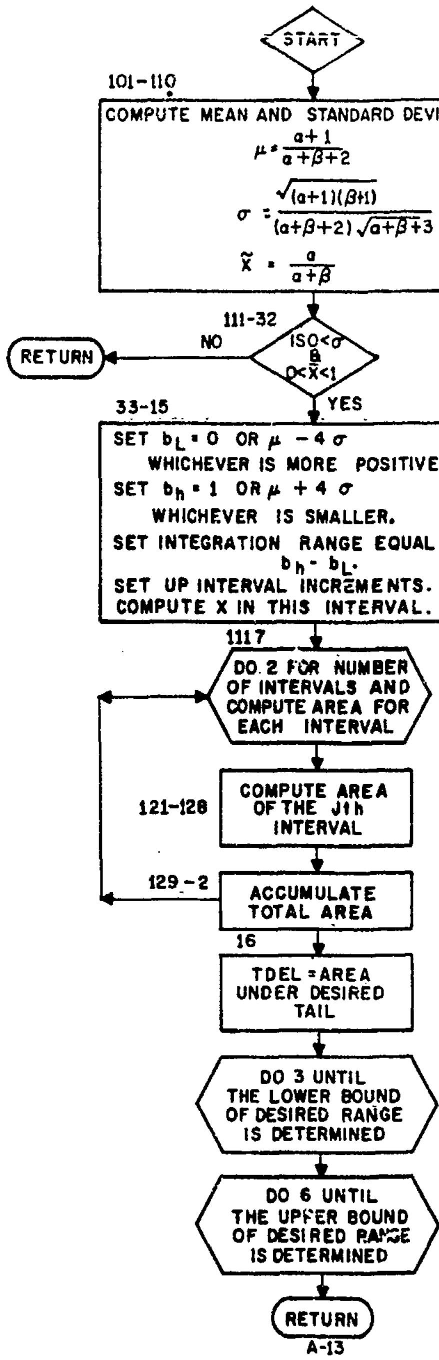



Figure 35. Program 111: Subroutine GASHUN

where

$$b_L = \max\left(\emptyset, \mu_N - 4\sigma_N\right)$$

$$b_H = \min\ 1, \left(\mu_N + 4\sigma_N\right)$$

$$\tilde{X} = \frac{\alpha}{\alpha+\beta}$$

Simpson's rule is used to evaluate the area. The formula for the area under the curve over a particular segment is:

$$A = \frac{1}{3} h\,[Y(1) + 4Y(2) + 2Y(3) + 4Y(4) + Y(5)]\ ,$$

where the Y's are five Y-values taken at five points within this segment on the X-axis.

The area for each of NINT segments (NINT is an input) is computed and added to the total of all previous ones to get the total area. Then the segments are accumulated until their total is greater than A times the total area (A is an input). The X of the last segment is the lower bound (XNCL) for the tail desired. The upper bound (XNCH) is determined similarly.

## REFERENCES

1. L. J. Savage, The Foundations of Statistics, New York, John Wiley and Sons, Inc., 1954.

2. R. Schlaiffer, Probability and Statistics for Business Decisions, New York, McGraw-Hill Book Company, Inc., 1959.

3. H. Raiffa and R. Schlaiffer, Applied Statistical Decision Theory, Boston, Harvard Graduate School of Business Administration, Division of Research, 1961.

4. C. J. Grayson, Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators, Boston, Harvard Graduate School of Business Administration, Division of Research, 1960.

5. M. V. Jones, "A Generalized Cost Structure for Electronic Systems," Bedford, Massachusetts, The MITRE Corporation, TM-3299, 1962.

6. J. E. Kelley, "Critical Path Planning and Scheduling," ORSA Journal, 1960.

7. Pearson and Harley, Biometrika Tables for Statisticians, University Press, 1956.

8. M. Zelen, "Bounds on a Distribution Function That Are Functions of Moments to Order Four," J. Res. Nat. Bur. Stan., 53, 1954.

9. E. Berman, "Toward a New Weapon System Analysis," The Rand Corporation, P1493, September 1958.

10. S. D. Fields et al., "Measuring the Reliability in Cost Analysis," TEMPO (RM #62-TMP-83), January 1963.

| | |
|---|---|
| The MITRE Corporation<br>Bedford, Massachusetts | 2b GROUP<br>Unclassified |

| | | |
|---|---|---|
| 3 REPORT TITLE<br>A COMPUTERIZED TECHNIQUE TO EXPRESS UNCERTAINTY IN ADVANCED SYSTEM COST ESTIMATES | | |
| 4. DESCRIPTIVE NOTES (Type of report and inclusive dates)<br>N/A | | |
| 5. AUTHOR(S) (Last name, first name, initial)<br>Sobel, S. A. | | |
| 6. REPORT DATE<br>November 1965 | 7a. TOTAL NO. OF PAGES<br>116 | 7b. NO. OF REFS<br>10 |
| 8a. CONTRACT OR GRANT NO.<br>AF 19(628)-2390<br>b. PROJECT NO.<br>850<br>c.<br>d. | 9a. ORIGINATOR'S REPORT NUMBER(S)<br>ESD-TR-65-79<br>9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)<br>TM-03728 | |
| 10. AVAILABILITY/LIMITATION NOTICES<br>Qualified requestors may obtain from DDC.<br>DDC release to CFSTI authorized. | | |
| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY<br>COMPTROLLER (Programs Division) Electronic Systems Division, L. G. Hanscom Field, Bedford, Massachusetts | |

13. ABSTRACT

The technique described presents a method to express uncertainty quantitatively in advanced system cost estimates. In particular, the technique suggests the employment of subjective probability distributions, which describe the uncertainty in each system element, to determine an approximate distribution for total system cost. A 7090 program has been written to perform the computational operations.

| | | | | | | |
|---|---|---|---|---|---|---|
| Probability<br>Distributions, Advanced System Costs<br><br>System Costs<br>Probability Distributions<br>Uncertainty<br><br>Computer Techniques<br>Probability Distribution, Determination<br>Uncertainty | | | | | | |

## INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

Security Classification